ISSN : 2456-3307

NEW HORIZON COLLEGE OF ENGINEERING

Autonomous College, Affiliated to VTU | Approved by AICTE New Delhi & UGC

50 Golden Jubilee YEARS OF EXCELLENCE IN EDUCATION

TECHNO SCIENCE ACADEMY
The International Open Access Publisher

IJSRCSEIT

Second National Conference on Internet of Things : Solution for Societal Needs

Organized by

Department of Information Science and Engineering,
New Horizon College of Engineering,
Kaverappa Layout,Near Marathahalli,
Bengaluru-560103 Karnataka, India

Volume 8, Issue 4, May-June-2021

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN COMPUTER SCIENCE, ENGINEERING AND INFORMATION TECHNOLOGY

Impact Factor Value 2020 = 6.135

Email: editor@ijsrcseit.com

Second National Conference on Internet of Things : Solution for Societal Needs

On

30th June 2021

Organized by

NEW HORIZON COLLEGE OF ENGINEERING

Autonomous College, Affiliated to VTU | Approved by AICTE New Delhi & UGC

Department of Information Science and Engineering, New Horizon College of Engineering, Near Marathahalli, Bengaluru-560103 Karnataka, India

In association with

International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307

Volume 8, Issue 4, May-June-2021

Published By

Techno Science Academy

The International Open Access Publisher

website : www.technoscienceacademy.com

CHIEF PATRON

Dr. Mohan Manghnani
Chairman
New Horizon Educational Institution, Bengaluru

PATRON

Dr. Manjunatha
Principal
New Horizon College of Engineering, Bengaluru

ADVISORY COMMITTEE

Dr. Amarjeet Singh
Prof & Dean- Academics, NHCE

Dr. K. GopalaKrishnan
Dean, Research and Development

CONFERENCE CHAIR

Dr. R. J. Anandhi
Prof & Head - ISE

CONFERENCE COORDINATORS

Dr. Arvind S. Kapse
Professor, ISE

Dr. P. Mangayarkarasi
Associate Professor, ISE

Dr. L. Srinivasan
Associate Professor, ISE

ABOUT NEW HORIZON COLLEGE OF ENGINEERING (NHCE)

New Horizon College of Engineering is an Autonomous college affiliated to Visvesvaraya Technological University (VTU), approved by the All India Council for Technical Education (AICTE) & University Grants Commission (UGC). It is accredited by NAAC with 'A' grade & National Board of Accreditation (NBA). It is one of the top engineering colleges in India as per NIRF rankings– 2018 and an ISO 9001:2008 certified institution. New Horizon College of Engineering is located in the heart of the IT capital of India, Bangalore. The college campus is situated in the IT corridor of Bangalore surrounded by MNCs and IT giants such as Intel, Accenture, Capegemini, ARM, Symphony, Wipro, Nokia, JP Morgan and Cisco to name a few.

ABOUT DEPARTMENT

Dept. of Information Science and Engineering at NHCE was established in the year of 2001 and offers B. E., M. Tech and Ph.D. programs and Accredited by National Board of Accreditation (NBA).

B.E. Information Science and Engineering

The four year B.E degree in ISE equip the students to meet day- to-day Technological advancements of the ever dynamic IT field through adept training on various subjects of curriculum of Information Science and Engineering and beyond. The department offers B.E program through autonomous scheme from the year 2015. The department has a total intake of 180 every year students with a very good team of highly qualified and talented faculty members including Professors, Associate Professors and Assistant Professors.

M. Tech

The department of Information Science and Engineering also offers 2 year M.Tech program in Cyber-Forensics and Information Security.

Ph. D.

The R&D Center has been established in the department in the year 2009 recognized by The Visvesvaraya Technology University (VTU)

ABOUT THE CONFERENCE

The National Conference on "Internet of Things: Opportunities for Society" (NCOIOT-2021) provides a common forum for deliberations, sharing of recent trends, advancements and research of Internet of Things in the areas of Science, Technology and Engineering. This conference will be a very good platform for academia, researchers, industry practitioners and technologists from all over the world to discuss and present recent advances and research outcomes in their respective fields.

The IoT is a highly emergent field in research and practice. It is difficult to find absolute and precise definitions of what constitutes the IoT. Moreover, it is impossible to limit the IoT to any specific technology. For instance, RFID technology is often associated with the IoT, as it permits users to assign a digital identity to a physical object (e.g., for the purposes of logistics). However, that does not mean that every application of RFID is part of the IoT. Rather, the IoT is largely defined by the manner in which devices are connected, thus permitting data to be shared, combined and used.

NCOIOT-2021 aims to provide an environment where the authors and participants can establish research relations and collaborations with various eminent academicians, research fellows, scientists from India and abroad.

CALL FOR PAPERS

We invite technical papers on, but not limited to the following domains.

- **Wireless networks for IoT and Cloud**
- **IoT-enabled home networks**
- **IoT and Smart cities**
- **Hybrid cloud infrastructure for IoT**
- **IoT based Application development**
- **Mathematical modeling for IoT**
- **Sensor Systems for IoT**
- **IoT applications in Building Automation**

CONTENTS

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Review on Classification of Techniques for IoT Based Home Security

Ajit Vikasrao Gaikwad[1], Dr. Arvind S. Kapse[2]
[1]Sr. Software Engineer, Xoriant Solutions Private Ltd, Pune, Maharashtra, India
[2]Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Internet of Things(IoT) refers to the infrastructure of connected physical devices that square measure connected through Arduino and multiple sensors and do wireless communication through web. Home security can bell be an awfully helpful application of IoT which we square measure exploitation it to make a security system for homes and industrial security purpose. The system can inform the owner concerning any unauthorized entry or whenever the door is opened by causation a notification to the user. Once the user gets the notification, he will take the necessary actions. This security system uses a microcontroller as Arduino UNO as associate degree interface between the elements of connected devices through the sensors that monitor the standing, a buzzer to sound the alarm and sends SMS, email to receiver if any suspicious activity happens.

Keywords — IoT, Arduino Uno, network sensors, embedded C, Wi-Fi, cloud, home security system, Sensor based Security System.

I. INTRODUCTION

There are lots of on-going projects on intruder detection systems with seismic sensors. The main purpose of these systems is to protect the border line from unauthorized entrance. There are many intrusion detection algorithms to achieve this goal. The main features that these algorithms differ from each other are; Detection range, False alarm rate, Power consumption, Noise reduction algorithms, Classifying the intruder The main disadvantage of these seismic sensors is that they can detect earth movements or wind as an intruder. Therefore, algorithms must be efficient. Some of the algorithms available are: Looking for the regular cadence of a typical human gait. [2]. Measuring the statistical distribution and detecting the extreme deviations from mean (Kurtosis) According these algorithms, noise and the detection range are inversely proportional. In the real world, it is assumed that the noise level is approximately medium, which means that intrusion detection range is approximately 20 meters. Applying Kurtosis in a high noise area with a specific threshold value can result to the misdetection of the intruder, because the noise level of the area is not always the same. [4] Another approach can be pursued. This approach is using strings of geophones for intrusion detection, and summation of the seismic signals. This contributes to the detection range. Even though all of the geophones have separate noises, only one geophone which is the closest to the walking person provides the main intruders signal.

II. WORKING PRINCIPLE

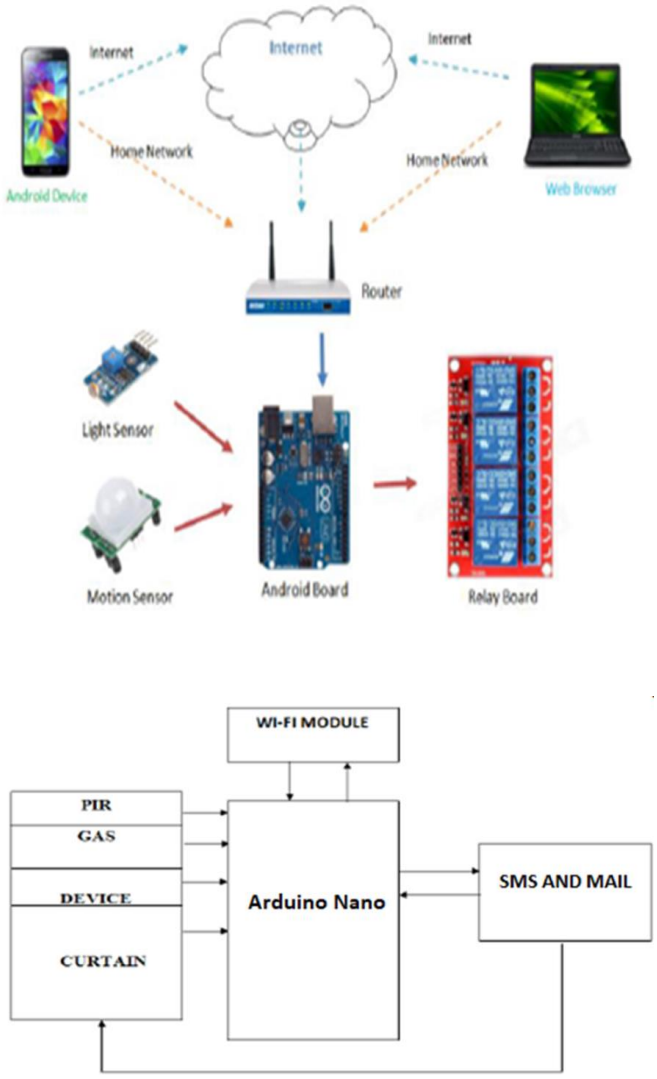


Fig1. Overview and Block Diagram of Proposed Project

III. PROPOSED SYSTEM

Our proposed system is an Arduino based home automation through with Arduino connected to a Wi-Fi and controlled via android app or a social media network. This system deals with the safety in home and smart home technologies, which will be cost efficient. Block Diagram of the proposed system is shown in Fig (1)

Arduino can sense the environment by receiving input from a range of sensors and might affect its environment via actuators. An analog temperature sensor is a chip that tells you what the ambient temperature is. The DHT11 might be a basic, ultra-low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and Hermiston to live the encircling air, and spits out a digital signal on the information pin (no analog input pins needed). It's fairly simple to use, but needed accurate timing to grab data. The only real downside of this sensor is you will be ready to only get new data from it once every 2 seconds, so when using our library, sensor readings are often up to 2 seconds old.

The Passive Infra-Red sensors allow one to sense motion, almost always and are used to detect whether a human has moved in or out of the sensors range. The PIR sensor could be a hydroelectric device that detects motion by measuring changes within the infrared level emitted by surrounding objects. This motion will be detected by checking for a high signal on a sign I/O pin.

They are small, economical, low-power, convenient and deteriorate. For that reason, they're commonly found in appliances and gadgets employed in homes or businesses. MQ6 may be a semiconductor type sensor, which might appropriately sense the presence of smoke, LPG, methane, butane, propane and other hydrocarbon combustible gases. The sensitive material in this sensor is tin-dioxide(SnO2). When it comes up-to-date with the gas to be monitored, the electric resistance of the sensor decreases; enabling the microcontroller to reply to the case. The sensitive material in this sensor is tin-dioxide(SnO2). When it comes up-to-date with the gas to be monitored, the electric resistance of the sensor decreases; enabling the microcontroller to reply to the case.

When it detects the concentration of combustible gas in the air it outputs its reading as an antilog voltage. The sensor can calculate concentrations of unstable gas of 300 to 10,000ppm. The sensor can operate at

temperatures from -20 to 50°C and consumes but 150 mA at 5 V. To allow connection for power plugs and switching of electrical load within the home, relay switches are used. The relay switches have capability to carry a maximum load of10A at 240V. This is sufficient to carry any household appliance as these devices do not draw much current. The Wi-Fi shield provides internet connectivity for the embedded micro web server which allows internet access and controls from a web application.

IV. LITERATURE REVIEWS OR SURVEY

Arduino Nano is a microcontroller board designed by Arduino.cc. The microcontroller used in the Arduino Nano is Atmega328, the same one as used in Arduino Nano. It has an extended range of applications and is a major microcontroller board because of its small size and flexibility. So, now let's have a look at its basic features:

Embedded System:

An embedded system is a computer system with a committed function within a bigger mechanical or electrical system, often with real-time computing constraints. It is embedded as part of a absolute device frequently including hardware and mechanical parts. Embedded systems run many devices in regular use today. 98 percent of all microprocessors are produced as elements of embedded systems. Examples of properties of typically embedded computers compared with general-purpose counterparts are low power consumption, small size, rugged operating ranges, and low per-unit cost. This comes at the worth of limited processing resources, which make them significantly tougher to program and to interact with. However, by building intellect mechanisms on top of the hardware, taking edge of possible existing sensors and also the extant of a network of embedded units, one can both optimally manage available resources at the unit and network levels also as provide augmented functions, well beyond those available. as an example, intelligent techniques may be designed to manage power consumption of embedded system.

An Embedded system could be a combination of hardware and software. like any electronic system, this technique requires a hardware platform which is constructed with a microprocessor or microcontroller. The Embedded system hardware includes elements like user interface, Input/output interfaces, display and memory, etc. Usually, an embedded system contains power supply, processor, memory, timers, serial transmission ports and system application affiliated circuits.

Embedded Systems are classified into three types supported the performance of the microcontroller like

1. **Small scale embedded systems**
2. **Medium scale embedded systems**
3. **Sophisticated embedded systems**

- **Small Scale Embedded Systems**

These types of embedded systems are designed with a single 8 or 16-bit microcontroller, that may even be activated by a battery. For developing embedded software for small scale embedded systems, the foremost programming tools are an editor, assembler, cross assembler and integrated development environment (IDE).

- **Medium Scale Embedded Systems**

These kinds of embedded systems design with one or 16 or 32-bit microcontroller, RISCs or DSPs. These kinds of embedded systems have both hardware and software complexities. For developing embedded software for medium scale embedded systems, the main programming tools are C, C++, JAVA, Visual C++, RTOS, debugger, source code engineering tool, simulator and IDE.

- **Sophisticated embedded Systems**

These sorts of embedded systems have enormous hardware and software complexities, which will need ASIPs, IPs, PLAs, scalable or configurable processors. they're used for cutting-edge applications that require hardware and software Co-design and components which need to assemble within the final system.

- **Embedded C**

Embedded C is a set of language extensions for the C Programming language by the C Standards committee to address commonality issues that exist between C extensions for different embedded systems. Traditionally, embedded C programming needs nonstandard extensions to the C language in order to support tropical features such as fixed-point arithmetic, multiple distinct memory banks, and basic I/O operations.

In 2008, the C Standards Committee increased the C language to address these issues by providing a common quality for all applications to adhere to. It includes a number of characteristics not available in normal C, such as, fixed-point arithmetic, named address spaces, and basic I/O hardware addressing.

Embedded C uses most of the syntax and semantics of merit C, e.g., main () function, variable definition, datatype declaration, conditional statements (if, switch, case), loops (while, for), functions, arrays and strings, structures and union, bit operations, macros, etc.

V. EXISTING SYSTEM

Gill et al. (2009) describes network enabled digital technology is rapidly established in the home automation. For the intent of home automation this technology introduces new and existing opportunities to increase the connectivity of the devices. The remote-control technology is quickly synchronizing with the growth of Internet.

Upadhyay et. Al. (2016) proposed a Home Indoor Positioning System (HIPS), gives location of mobile devices like smart phones and location comprised of IoT applications. This paper induces home indoor placed system using Wi-Fi signals. In suggested system an intelligent mobile robot automatically builds radio maps for the system.

Shetel and Agarwal (2016) describes in their paper that IoT authorize internet connectivity for all types of devices and physical objects in real time system. The virtualization of this system allows to perform project without direct physical synchronization among the devices. The IoT enables to run multiple jobs without any restriction of distances with the help of intelligent devices and high-speed network.

Lee et al. (2017) describes in their paper the web of physical objects is Internet of Thing which consist the embedded technology helping in constructing machine to machine or man to machine communication. This paper provides a dynamic data sheet about the city territory parameters taken from the stand-alone system.

Chou et al. (2017) describes in their paper a home automated system has remote controlled operation. This paper describes about the problem on their setup, finding out the multiple solutions through different network technologies and trying to improve the use of these system. The Home Automation System (HAS) needs heterogeneous, an eternal and transitive computing environment's careful study to develop the suitable HAS.

Kamal et al. (2017) describes in their paper how this paper used Raspberry Pi as the network gateway. This paper uses MQTT (Message Queuing Telemetry Transport) protocol for sending and receiving the data. All the sensors used in this paper is been managed by the web page executing the Access Control List (ACL) for providing encryption method for the secure transaction of the data's. This paper

makes use of multiple sensors, wired and wireless, are connected with the Raspberry Pi.

Sahadevan et al. (2017) describes in their paper how the Internet of Things is surprisingly affecting the attention of customer and the enterprise electronics market quickly implementing in home automation, smart cities, automated industries, etc. To construct these applications many efficient and low cost sensors are available in the market for the developers. The server side is taking caution of computational work whereas the client side is taking caution of sensor actuator work. This needs robust networking infrastructure over the world. This paper put forward an offline online asynchronous communication course of action for Internet of Things application where Message Queuing Telemetry Transport (MQTT) protocol is used. This paper has conveyed out a portable device system on Intel Galileo which exhibits the chance of such a system without come to an understanding the functionality.

VI. CONCLUSION

In this way as a part of home security purpose it become every essential to deal with home security parameters associated with it. In the proposed system it was cleared that whenever we need a physical parameter when we are not at home this system helps to maintain the security essentials by making use of some Arduino circuit and integrating it with IoT related features. It helps to secure the parameter of your house and maintained the essential need.

VII. REFERENCES

[1]. K. Ashton, That ―Internet of Things‖ Thing, RFiD Journal. (2009).

[2]. H. Sundmaeker, P. Guillemin, P. Friess, S. Woelfflé, Vision and challenges for realising the Internet of Things, Cluster of European Research Projects on the Internet of Things - CERP IoT, 2010.

[3]. J. Buckley, ed., The Internet of Things: From RFID to the Next-Generation Pervasive Networked Systems, Auerbach Publications, New York, 2006.

[4]. M. Weiser, R. Gold, The origins of ubiquitous computing research at PARC in the late 1980s, IBM Systems Journal. (1999).

[5]. Y. Rogers, moving on from weiser's vision of calm computing: Engaging ubicomp experiences, UbiComp 2006: Ubiquitous Computing. (2006).

[6]. R. Caceres, A. Friday, Ubicomp Systems at 20: Progress, Opportunities, and Challenges, IEEE Pervasive Computing 11 (2012) 14–21.

[7]. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless Sensor Networks: A Survey, Computer Networks 38 (2002) 393–422.

[8]. L. Atzori, A. Iera, G. Morabito, The Internet of Things: A survey, Computer Networks 54 (2010) 2787–2805.

[9]. J. Belissent, Getting Clever About Smart Cities: New Opportunities Require New Business Models, Forrester Research, 2010.

[10]. R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems 25 (2009) 599–616.

[11]. S. Tilak, N. Abu-Ghazaleh, W. Heinzelman, A taxonomy of wireless micro-sensor network models, Acm Mobile Computing and Communications Review. 6 (2002) 28–36.

[12]. http://rcciit.org/students_projects/projects/aeie/2018/GR10.pdf

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Modeling For Multicore System Simulator for Computer Architecture

Mr. Shirish Pattalwar[1], Dr. Vilas Thakare[2]
[1]Research Scholar, SGB Amravati University, Amravati, Maharashtra, India
[2]Professor & Head, Department of Computer Science, SGB Amravati University, Amravati, Maharashtra, India

ABSTRACT

This research discusses the various issues regarding the accurate and fast and automated system architecture which gives specific information about the various parameters and there effects on the simulation of the structure for the efficient processing of the system modeling. As there is a great demand of the simulation of the system architecture this research gives the better idea for the simulation and various components involved and how the process is followed superior quality of design and development components regarding the efficient utilization of the multicore processor. This research also discusses the various components like basic structure for simulation and for the efficient operation of the system using the various components and parameters which are closely related to each other. The detail analysis of these parameters is also done which are so intensely attached to each other that they may affect each other.

Keywords - Multi-core x86 CPU simulator; Emulator; Full- System simulator; Heterogeneous Multi-core systems; Processor Modeling.

I. INTRODUCTION

Now a days there is a great demand of the high end, fast and versatile devices which involves the high end processor and which also leads to different kinds of applications such as hard real time and soft real time. Any processor system before being implemented practically needs much iteration of up-gradations through simulation. The hard real time processors are those in which the deadline for the task assigned has to be completed within the specific and accurate timing constraints. Hence there is a great requirement for the high end processor and the cost of such processor design is very high. As it involves the number of critical issues which includes the design, development and implementation of such high end processors.

Hence before the actually implementing the hardware in to the hard core processor the various parameters regarding the processor must be studied and analyzed for the proper operation of the system and the overall functionality of the system architecture must be understood for the accurate functionality. Thus there is a great demand of such design which will fulfill all the design requirements which are in continuous demand for the application like military applications.

Another important aspect in the design and development of such system architecture involves the proper memory management. As all the data or information on which the processor is going to

process is present in the memory of the system. In other words we can say that there is a great demand of memory which can be less in cost and saves the most of the necessary information in the processor chip. As by default all the data is stored in the secondary memory of the system. It has to be brought in the main memory or the primary memory for the operation which is then must transferred to the small memory in the processor chip. This small memory has to be on the processor chip. If the data or information which is to be processed by the processor is present in the small memory present on the processor chip then the operation the processor can perform is efficient, fast, accurate and without any delay, otherwise the time required to get the data into the processor is involved in the operation of the processor is going to perform.

Therefore the proper selection of the memory requirement has to be judged accordingly which will satisfy the needs of the application where the processor is to be implemented. The processor has to deal with the various resources and input output devices and needs to be interface with the other integrated devices. The resources may involve the chipset, Dynamic RAM, network interface cards and peripheral devices in addition to this there are number of devices the processor has to deal with.

Thus there is a great requirement for the accurate simulation models which can fulfill the system modeling for the high end processing also with respect to the kind of application involved in the designing phase. The use of single and multicore processors are frequently done in the design, development and implementation of the system architecture which can be very useful for the user for the performing the various operation regarding following the strict deadline of the task which is assigned to the processor.

There is another parameter which plays a very important role in the design, development and implementation. It is the cost of manufacturing. The designer or the architect of the system must always consider this factor as a high priority which is very essential.

In the development of the system architecture the most essential and important dimension is the use of the proper simulation tool. If the simulation tool is easy to use the designer can develop and use the tool very efficiently and make the system faster and efficient. The simulation tool is very essential for the systematic development and implementation of the parameter which can successfully follow all the steps for the proper operation and systemic mechanism of the development of the processor modeling. Thus the simulation model or simulation tool has to be realistic which involves all the parameters for the proper development of the system architecture.

Such a tool is useful for evaluating and developing products that will use current and emerging single and multicore x86 chips.

This research introduces another dimension in the development of the system architecture which is very useful for the number of application involving the hard real time requirements and soft real time requirements[1].

The other system may use different types of the simulator for the simulation of the various types of the applications involved in the production of the online codes for the IOT industry, such as there are number of system which automatically generates the codes depending upon the system input. Such types of the systems give the better performance with respect to the soft real times as there is not accurate deadline attached or needed for the application to run and hence there is more demand of such type of systems which are useful for the generation of code which are executable for the various applications in different domain.

The cost of such systems is very low as compared to the systems those require the accurate timeliness. If the application requires the accurate timeliness then,

such type of the systems also use the high end processor hence the cost involved is maximum[2].

To the best system architecture it has to go through the various types of stages for that purpose knowledge and use of proper simulation tool is very essential, System architecture is not aware of issues in the public domain, open source simulator that rivals the characteristics of different systems.

This research focuses on the various dimensions of the system implementation and their details regarding the evaluation of the performance of various issues of multi- threaded benchmarks.

II. BACKGROUND

Many research studies have been done on the various simulator designs also to get more perfect simulation results for various parameters.

The out-of-order processing for the core designed is used in various applications which may the issue of the simulation time. In other words it decreases the simulation time to a great extent hence there is a great demand of such type of the simulation tool[1] with the cost of the processing is increased significantly which may hamper the total manufacturing cost of the complete system architecture and evaluation of the performance in some of the application[7].

In order to reduce the cost of the simulation there are many types of the simulation tool which are used such as PTLsim. MPTLsim simulator which may significantly reduce the cost of the simulation and on the other hand the complete system architecture can be developed in the very less cost[3].

The paper is arranged in following manner:

Section I Introduction. Section II describes Background. After that Section III describes previous work done by various researchers in this domain. Section IV describes various existing methodologies. Section V describes analysis & discussion with base approach of proposed methodology. Section VI is approach for testing. Section VII is proposed methodology. Section VIII is outcome and result Section IX is conclusion. Section X is future scope of this paper.

III. PREVIOUS WORK DONE

In research literature, many simulation technologies have been studied to provide various aspects of this research and improve this technology with maximum output.

The proposed system architecture is useful, integrated and accurate simulation frameworks develop for a broad range of applications[2]. This framework consists of various elements like a internal memory management unit, main CPU emulator and different IO devices for performing input and output, and chipsets operations.

This research gives a very high-level approach of various elements of the proposed model along with the different CPU simulation tools and framework[3]. This work postulates that the coarse-grained migration in existing heterogeneous processor designs may limits their effectiveness and energy savings. This system provides the Composite Cores for different aspects for the emulation of the system components associated with the heart of the computation and different features for the best system development[5][6]. Hence the system discussed can achieve both energy efficiency and high performance. In such work, the system promotes a dual nature of Composite Core using the components for better simulation.

IV. EXISTING METHODOLOGIES

Many techniques have been implemented over the last several decades on Composite Cores, an architecture that brings the concept of heterogeneity into a single core. The performance and energy

implications of several architectural designs of a Composite Core with the goal of maximizing both energy savings and physical layout have been analyzed. System proposed the addition of a small L0 filter cache [8] for the little mEngine, as well as evaluates the effects of various migration techniques. System evaluates proposed Composite Core architecture with cycle accurate full system simulations and integrated power models. Overall, a Composite Core can map an average of 25 percent of the dynamic execution to the little mEngine and reduce energy by 21 percent while bounding performance degradation to at most 5 percent.

gem5-gpu routes most memory accesses through Ruby, which is a highly configurable memory system in gem5. By doing this, it is able to simulate many system configurations, ranging from a system with coherent caches and a single virtual address space across the CPU and GPU to a system that maintains separate GPU and CPU physical address spaces[9].

Single and multicore processors implementing the x86 instruction set architecture (ISA) are deployed within many computing platforms today, starting from high-end servers to desktops and ultimately down to mobile devices (using the Intel Atom and its announced successors), including potential new products that target the smart phone market segment and beyond. The one clear advantage of using the x86 processors in the full range of the product spectrum is to facilitate the rapid deployment of the wide variety of x86 application binaries. It is thus important to have a full system simulation tool that incorporates realistic simulation models for other systems level components such as the chipset, DRAM, network interface cards and peripheral devices in addition to accurate simulation models for single and multicore processors implementing the x86 ISA[30]. Such a tool is useful for evaluating and developing products that will use current and emerging single and multicore x86 chips. MARSS uses a cycle-accurate simulation models for out-of order and in-order single core and multicore CPUs implementing the x86 ISA[10].

The researchers studied different techniques for the better simulation and composition of various units for the better design of the system architecture and also studied and analyzed various parameters regarding the various situation concerning to the various critical situation in which the system has to work for the various application domains.

V. ANALYSIS AND DISCUSSION

The proposed methodology uses advanced simulation tools, which are very useful in judging the performance of the proposed system architecture which is a integration of various processor and various advanced computational units for the combined use of the logical, arithmetic and logical functions of the system for performing the various complicated task associated with computation. Thus the system will provide accurate simulation and analysis of the critical issues occurred in the process of the simulation for better simulation experience, modeling the decomposition is done using the Virtual Machine Framework and is extensively modified to realize some important features of the proposed system architecture. This framework is extremely useful for simulation/emulation and extensive changes are done to make the system applicable to various application domains. The proposed simulation techniques and models are really useful in multicore microprocessors with coherent memory design, Dynamic memory systems and various interconnections including on-chip interconnections and various instructions for execution.

A. Base Approaches for Methodology

There are various priorities and privileges assigned to the different type of task associated with the simulation of the processes that the system architecture is going to perform or in other wards

system is going to perform thus the situation becomes very simple that the task having higher priority or privilege will be executed first depending upon the various resources the process is going to used. There are also many instances of each resource that a system may have and must be allocated in such a way that there should be optimum use of the resources. Hence the priority of the process will help to resolve the conflict if any which is essentially useful for the system in which there are more number of processes and resources are limited. Another approach the system must allow is to use a specific amount of time required for each process and each process is allocated a specific amount of time which is called as quantum time. After completion of this time the resources is taken away from the process and is allocated to another process. This ideas or approach is very useful where there is limited amount of cpu time and that is to be distributed among the various processes associated with the system.

All above discussed scenario is simulated among the different tools for the system architecture simulation. The simulation tool which can perform best or which is of great use in the judging the performance of the system architecture modeled is very essential and useful for the actual implementation and therefore, the selection of the simulation tool for the multicore system plays a very essential and important role in the performance evaluation of the system architecture and design.

The basic concept behind the selection of the proper simulation is that the simulator tool should be efficient and effective while judging he performance of the overall system architecture proposed and there should be no errors in the modeling of the system architecture. The system architecture modeled must be accurately modeled by the simulation tool used.

There are various types of simulation tool available in the market, which are particularly working or focusing on different areas of the simulation, but the designer or architect of the system must use some strategy for the selection of the simulation tool among the various types of the tool available for the simulation depending upon the domain of the applications and the need of the applications. Thus the selection of the simulation for the multicore system has to be done very carefully to enhance the performance of the system in order to achieve the system designed goals.

Another factor which is very important while selecting the simulation tool is the simulation speed. In contrast with the performance judging capacity of the simulation tool, system designer must also concentrate on the various issues of the simulator tool like simulation speed. If the simulation speed is too slow then such simulation is of no use and leads to unnecessary delay in the simulation which the system may not tolerate. In such a case high speed system simulator are very effective and efficient.

In addition to this, the system architecture must use stack registers and allows direct access to most of such registers to improve the speed of the operation. The system must use add on support for instructions to be executed and register accesses, allowing users to use pre-compiled applications and library functions that utilize system registers.

Communication between Simulator and applications provides a unique feature to system architecture to send and receive instruction from the system simulator and the application where the system to be implemented. This also has modified mechanism to reflect the simulator cache configuration like memory size, memory line, and shared/private memory to be used. Thus the application enables simulator features to optimize.

VI. APPROACH FOR TESTING

The proposed system architecture uses a event-based memory management for system simulator to increase the simulation speed according to availability of the memory in the system. Such

simulation models memory and different controllers and management protocols for the simulation, on-chip interconnections and a simple DRAM controller. The proposed system has to be verified and validated against the different characteristics of the desired simulated system which helps in the monitoring the performance of the proposed system against the different parameters which are associated with the various issues of the system simulation process. These parameters are closely linked to each other and may affect the system in different dimension which is finally essential for judging the overall performance of the system architecture.

Hence the parameters for the simulation must be carefully analyzed for their effect on the system simulation.

The system must maintain event queues for the purpose of the system simulation which is very essential for the testing the various parts of the simulation process. The event queues are modeled in such a way that they does not specify the event compilation order specified within particular simulated cycle.

This proposed system simulation model simulates non- uniform memory accesses and delay in modeling unlimited size of memory accesses within a controller. The system also gives best three models of memory: simple write-back memory, simple write-through memory, and coherent memory implemented using the specified protocol design. Figure indicates a specified flow of diagram of the logic in a very simple memory module. All the different edges between the various processes blocks may indicate quantum of delay to be simulated between various events executed.

VII. PROPOSED METHODOLOGY

Model Refinement through Compact Trace Transformations (CET) methodology is implemented on the basis of various components and related events like coherent caches, interconnections, chipsets, memory, IO, executions and its timings. This stage of simulation is a beginning step of simulation that uses Trace driven approach combined with flow based approach and full system simulation. To begin with it first focuses on level of abstractions.

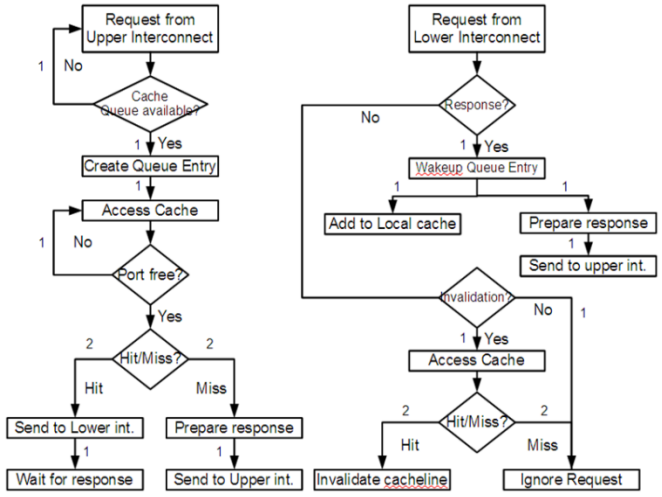


Fig. 1 Flow diagram of CRM logic with delays

A. Level of abstraction

Level of abstraction relates to every subcomponent of the system. So, Refining architecture model components in full architecture system level simulation framework requires that the application events driving these components should also be refined to match the architectural detail.

B. Trace transformations & CET

Refinement of part of system application events is denoted using trace transformations in which the left-hand side contains the coarse-grained application events that need to be refined and the right-hand side the resulting part of architecture-level events.

Furthermore, "!" symbols in trace transformations denote the "followed by" ordering relation. The R & W trace transformations are given below that refines Read and Write application events such that the synchronizations are separated from actual data transfers.

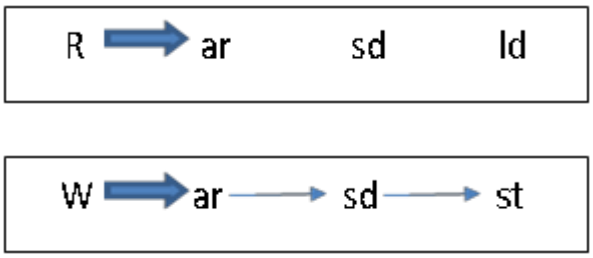


Fig. 2 Flow of shows the logic with delays in processing.

Where, events ar is activate data *
sd is signal data > ld is load data !
st is store data !

The events marked with * refer to synchronizations, while those marked with > refer to data transmissions & ! refers to followed by.

C. Event Refinement Using Dataflow Graphs

In a system, Synchronous Data Flow (SDF) and Integer controlled Data Flow (IDF) actors are deployed to realize trace transformations. As it is shown in this section, the SDF actors perform the actual event refinement while dynamic IDF actors are utilized to model repetitions and branching conditions that are present in the application code. In addition, IDF actors also be used to achieve less complicated (in terms of the number of actors and channels) dataflow graphs. It is assumed the fact that SDF is a subset of IDF.

VIII. OUTCOME and RESULTS

Model for Architecture refinement through trace transformations CET overloads:

CET = n × IDFt

IDFt = SDF1t + SDF2t + SDF3t + …….n

SDFt = Rt + Et + Wt

Rt = tar + tsd + tld × Bt Wt = tar + tsd + tst × Bt

Where, t indicates the time required for every event and actor B indicates block operations with n = no of actors

CET Overloads with Et=0.25, n=1:

TABLE I. SDFT & CET DETERMINATION (FOR ET = 0.25, N = 1)

Sr. No.	tar	tsd	tld	tst	Bt	Rt	Wt	SDFt IDFt CET
1	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.0
2	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.0
3	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.0

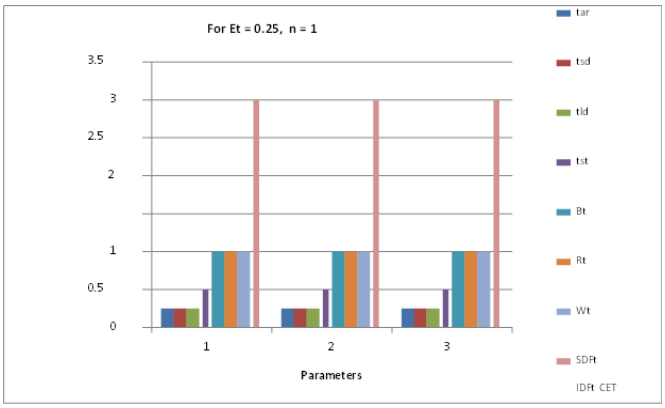


Fig. 3 Chart for Model for Architecture Refinement through Trace Transformation.

CET Overloads with Et=0.5, n=1:

TABLE II. SDFT & CET DETERMINATION (FOR ET = 0.5, N = 1)

Sr. No.	tar	tsd	tld	tst	Bt	Rt	Wt	SDFt IDFt CET
1	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.25
2	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.25
3	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.25

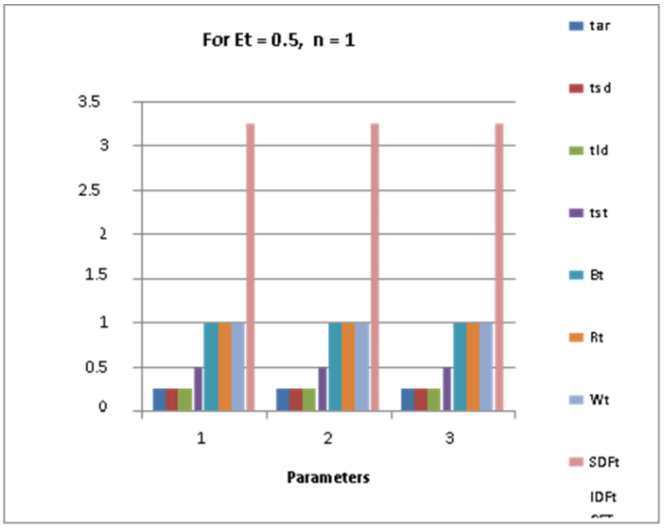


Fig. 4 Chart for Model for Architecture Refinement through Trace Transformation (CET Overloads with Et=0.5, n=1)

CET Overloads with Et=0.75, n=1:

TABLE III. SDFT & CET DETERMINATION (FOR ET = 0.75, N = 1)

Sr. No.	tar	tsd	tld	tst	Bt	Rt	Wt	SDFt IDFt CET
1	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.5
2	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.5
3	0.25	0.25	0.25	0.5	1.0	1.0	1.0	3.5

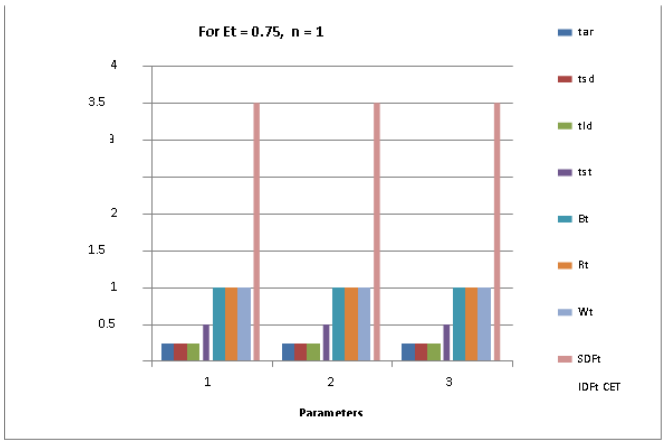


Fig. 5 Chart for Model for Architecture Refinement Through Trace Transformation (CET Overloads with Et=0.75, n=1)

IX. CONCLUSION

This research analyzed the various factors for the better simulation techniques for the best system architecture development scenario considering the various parameters such as CPU time, memory, performance, user convenience and efficiency and effectiveness with respect to different platforms specially for the multicore system development and provides the unique model for the simulation of the complete system development with high efficiency and less cost.

X. FUTURE SCOPE

It is expected that the continuous research and development will eventually result in new design of multicore and multithread processor simulator with more simulation speed and accuracy. These strategies will also improve the effectiveness and efficiency of simulation.

XI. REFERENCES

[1]. B. Christopher, P. Vaidya, and J. L. Jaehwan, “An XML-Based ADL Framework for Automatic Generation of Multithreaded Computer Architecture Simulators”, IEEE Computer Architecture Letters, Vol: 8 Issue No: 1, 13-16, January 2009.

[2]. H. Morteza, “Flow-Based Simulation Methodology”, IEEE Computer Architecture Letter Vol: 17, Issue No: 1, 51-54, January 2018.

[3]. J. K.Archibald, “An Innovative Simulation Approach for Labs in Computer Architecture”, 3rd ASEE/IEEE Frontiers in Education Conference, 19-24, November 9. 2002.

[4]. G. Braun, A. Nohl, A. Hoffmann,O. Schliebusch, R. Leupers, and H. Meyr, “A Universal Technique for Fast and Flexible Instruction-Set Architecture Simulation”, IEEE Transaction on

Computer Aided Design of Integrated Circuits and Systems, Vols.: 23 No.: 12, 1625- 1639, December 2004.

[5]. E. Schneider and H. J. Wunderlich, "SWIFT: Switch-Level Fault Simulation on GPUs", IEEE Transaction on Computer Aided Design of Integrated Circuits and Systems, Vol. 38, No. 1, 122-135, January 2019.

[6]. A. Zjajo, M. Eijk, R. Leuken, C. Strydis, "A Real-Time Reconfigurable Multichip Architecture for Large-Scale Biophysically Accurate Neuron Simulation", IEEE Transactions on Biomedical Circuits and Systems, Vols. 12 No.: 2, 326-337, April 2018.

[7]. S. Lee and W. W. Ro, "Parallel GPU Architecture Simulation Framework Exploiting Architectural-Level Parallelism with Timing Error Prediction", IEEE Transactions on Computers, Vols. 65, No. 4, 1253-1265, April 2016.

[8]. A. Lukefahr, S. Padmanabha, R. Das, F. M. Sleiman, R. G. Dreslinski, T. F. Wenisch, and S. Mahlke, "Exploring Fine-Grained Heterogeneity with Composite Cores", IEEE Transactions on Computers, Vols. 65 No. 2, 535-547, February 2016.

[9]. J. Power, J. Hestness, M. S. Orr, M. D. Hill, and D. A. Wood, "gem5- gpu: A Heterogeneous CPU-GPU Simulator", IEEE Computer Architecture Letters, Vols. 14, No. 1, 34-36, June 2015.

[10]. A. Patel, F. Afram, S. Chen, and K. Ghose, "MARSS: A Full System Simulator for Multicore x86 CPUs", DAC'11, June 5-10, 2011, San Diego, California, USA ,1050-1055, June 5-10, 2011.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Design of an IoT-Based Traffic Management System for Smart Roads

Deepa R[1]*, Gnandesigan R[1], Ranjith D[1], Siva Priya S[1], Sakthipriya S A[1], Tharunika N[1]
[1]Department of Electronics & Instrumentation Engineering, Bannari Amman Institute of Technology Sathyamangalam, Tamil Nadu, India

ABSTRACT

The primary objective of this design is to prevent traffic accidents. The great deal of research strongly indicated that the most of incidents happen in specific locations, such as school zones, restaurants, and so on. The proposed system's design is to avoid accidents in nearby areas by alerting or warning via Wi-Fi. A mobile notification is transmitted through the Wi-Fi link if the zone is fewer than 110 yards from wagon. In reply to this warning, the motorist slows the wagon's speed. A Wireless Fidelity warning will be sent to a cell phone if a school region is within 110 yards of it. After receiving alert from the Wireless Fidelity range to the driver's notification, the vehicle's speed is reduced. It provides message warnings to mobile devices within 100 metres of a Wi-Fi hotspot, as well as changes in weather parameters. The driver slows the vehicle's speed in response to this command. This sign contributes to a reduction in traffic based accidents.

This system's controlled power supply uses 5V and 500mA, three voltage terminal regulators (7805) were utilised for voltage power and A full-wave bridge rectifier is used to rectify the 230/12V transformer secondary ac output. We have used Wi-Fi AT89S52 switches and a Wireless Fidelity module to connect the controller to the Internet of Things, allowing it to be operated from anywhere through IoT.

Keywords : Traffic management system, Internet of things (IoT), Wi – Fi, Vehicles

I. INTRODUCTION

In order to accomplish traffic operations service provider aims, automatic traffic regulation systems include traffic signals. Intersectional traffic lights, additional network of communications that connects them as well as a central hub of computer or network of machines to supervise an operation are all part of these systems. Coordination is crucial and it may be enforced using a variety of mechanisms, including time-based and hardwired communication systems. Agency-wide traffic signal the creation of agreements on data interchange and traffic signal regulation are examples of collaboration. As a result, forming conventional and non- conventional consensus between jurisdictions to transmit traffic management and signal operation skills is a significant policy component of indicating the Traffic Signal. A traffic sign arrangement's main aim is to present drivers with a favourable signal period. The programme includes tools to help the developer accomplish that goal to the best of their abilities Adjustments and corrections in the signal of the systems are used to offer power. Those are the most important control features. They have access to the junction signal controller for care and upkeep. If the connection is complete and secure, the operator will become high effective and efficient. Several monitoring functions

are now accessible, including traffic detection, surveillance footage, and signal traffic management. These also offer better efficient traffic management techniques, such as efforts to adapt traffic in a predictable and adaptive manner.

Many metropolitan areas are experiencing issues due to population growth. This has resulted in a rise in automobile density; As a result, there is traffic congestion. Vehicle speeds are slowed during rush periods and during other times due to the change in traffic density. The current infrastructure can only regulate traffic jams with a limited number of resources.

IOT has been used to provide real-time traffic density control in order to manage traffic flow. It helps with traffic flow control, traffic shifting efficiency, and excessive avoidance [1]. These capabilities have been added to the website to show the current traffic status, allowing visitors to get an early update and avoid traffic jams by taking an alternate route. In the event of an emergency, vehicles can receive early access to their destination.

In an environment where native species are becoming rare and leftover animals are being driven there by vehicles when using roadways, IoT customs a concept to reduce accidents between wild life species and wagons on major roadways in protected regions [2]. Because electric fence poses a significant threat to animal life, our article proposes a wonderful alternative: providing notifications via smart phones and large LED displays strategically positioned across important areas where animals are spotted crossing highways. The approaches of observing the movement with certain devices and visual perception using machine learning techniques are applied.

In the United arab Emirates (UAE), vehicle violent incidents with regard to animals (especially camels) are growing increasing regularly, result in economic losses, animal deaths. Annually, there is significant loss in species. As a consequence, design and implementing a dependable plan for detecting wild life species and alerting drivers in the UAE has become a critical necessity. Although some such methods have been used in some other parts of the world [3]. The goal of this study is to create a suitable solution to avoid the occurrence of accident with wild life species by using the Internet of Things reliable r system that is cost-effective. The wild life species identification and driver safety systems seem to be the two key components of the system. Small level of data flow rate is required while uploading to the cloud. As a result, the chances of data traffic are minimal. It can also be employed at night, when most animal based vehicle accidents happen.

The wildlife populations can be severely impacted by road fatalities. Few studies, however, have looked at the effectiveness of strategies for reducing wildlife-vehicle collisions (WVCs). We tested highway avoidance barrier to reduce the collision between wagons and wild life along the path segments of the Trans-Canada main road in Banff National Park, Alberta (phases1, 2, and 3A). We studied onslaught of wild life species on the barried right-of-way from 1981 to 1999 [4]. We discovered that WVCs were not dispersed randomly after netting and were related to and close to fence ends. Wildlife-vehicle collisions are also most common within 1 km of barrier ends, though access to key drainages likely influenced collision location as well.

In all metropolitan cities, traffic congestion is a concern, particularly in the downtown areas. Normal civilizations can be transformed into smart cities with new technologies (ICT) by utilising data and transmission technologies [5]. The Internet of Things (IoT) paradigm has the capacity to play a big role in the evolution of Cities with smart capabilities. This is an Internet- of-Things-based traffic solution provider for smart cities, in which road traffic can be actively controlled by local traffic cops using their mobile phones or constantly observed and managed over the Internet.

Using the Internet of Things, information, and insights, a true transportation planning system is presented (TMS). Increased car numbers result in a plenty of problems, including wasted time and fuel, Contamination of the air and sound, and even mortality from stranded external crises wagons. Ultrasonic sensors are used to measure traffic density [6]. The system controller employs a traffic management algorithm to alter traffic signal timing after evaluating data and transfers information to the cloud host via a Wireless Fidelity device. It can predict whether or not there will be road traffic at the crossroads. If an essential vehicle is detected, the intersection receives priority and a longer signal time. The world is changing at a breakneck pace, and it must continue to do so in order to progress. However, contemporary transportation fails to offer citizens with a smooth transportation system. Excessive traffic jams cause delays in getting to work or home, waste of gasoline, vehicle wear and tear, and even road rage among stressed and irritated drivers. We frequently observe people waiting in large lines to pay toll taxes. Vehicle owners face an additional headache in the form of parking. We have offered a solution to all of these challenges using the Internet of Things in order to realise the Smart City's goal. We've devised an algorithm for traffic congestion control and smart parking system [7]. In today's world of ever-increasing population, the Internet of Things (IOT) plays a critical role. Its uses include autonomous transportation, smart homes, smart cities, agriculture, and health care [8].

The Iot technology is a promising new technologies have the potential to make our society's infrastructure smarter and responsive to its users' needs. A good example of such framework is conveyance. So this paper investigates contemporary approaches and techniques for automatic vehicle traffic control in order to identify hotspots and ways for introducing IoT into the field. Existing approaches are examined for their benefits and limitations, as well as their effectiveness [9].

The basic purpose of the road safety system is to keep people safe on the roads. so that accidents caused by a driver's incompetence can be identified and fatalities avoided. As a result, the intelligent traffic structure proposed is favourable due to the Internet of Things' unique traits and capabilities.

II. METHODS

The proposed methodology will be split into two parts: one for the emitter and one for the recipient. The transmitter is depicted in the block diagram in figure 1, which is separated into numerous zones. Figure 2 shows how the RF transmitter will receive data from the switches and sensors in those specified zones. The information is emitted with help of RF transmitter, and the data obtained by the RF receiver and processed by the receiver's microcontroller. We can monitor everything in the website after it has been processed because the gadget is also connected to the Wi-Fi module.

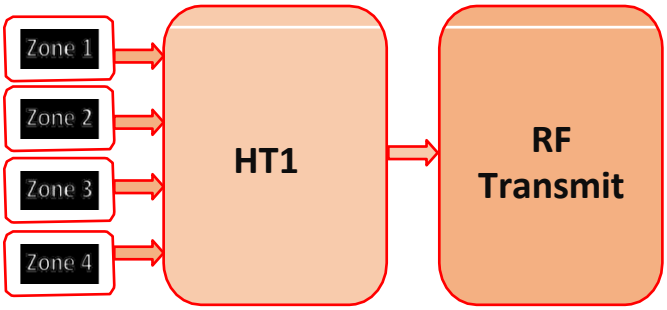


Figure 1 Transmitter Section

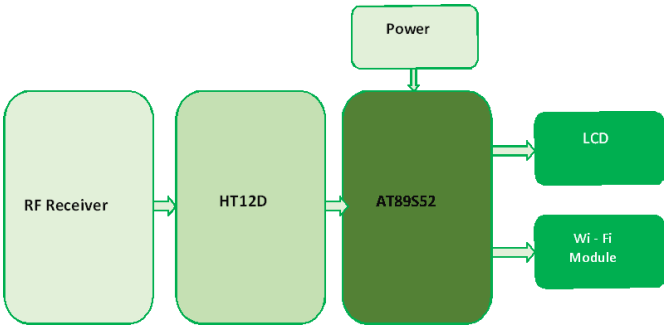


Figure 2 Receiver Section

III. MICROCONTROLLER – AT89S52

Figure 3 shows the AT89S52, an 8-bit, limited power, 8-bit CMOS microcontroller with an in- command Flash memory of 8 kilobytes. The machine uses Atmel's heavy non-volatile memory technology and follows the 80C51 pin and operation set, which is industry standard. You can use your computer or a escapees memory programmer to rewrite your program's memory utilising the Flash functionality. The Atmel AT89S52 is a sophisticated microcontroller that combines the configurable 8-bit CPU with a Flash configurable processor on a monolithic chip to produce a more versatile to time-saving solution for many applications in the field of embedded .AT89S52 contains 8 K bytes of Flash, 256 bytes of RAM, 32 I/O blocks, two data points, three 16-bit timers/counters, a six-vector double-speed serial interface, an on-chip oscillator, and a timer monitor. Regardless, AT89S52 has a null frequency static working concept that enables for various programming state saves. The Kernel is disabled in Idle State, and RAM, clock, and machine interrupt activities are performed. The OSC is frozen until the next break is completed, but the RAM contents are stored.

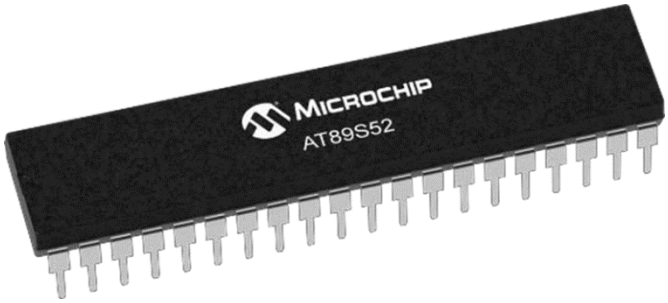


Figure 3 AT89S52 Microcontroller

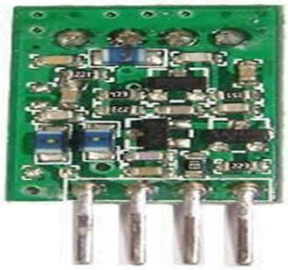


Figure 4 RF Transmitter

IV. WIRELESS RF TRANSMITTER

The STT-433, with its 433MHz RF transmitter portion illustrated in figure 4, is designed for small economy, high-range hand held applications. Because this transmitter runs on 1.5-12V, it's suitable for battery-powered applications. In order to achieve accurate frequency control and the best output, the transmitter uses a SAW-stabilized oscillator. Controlling output power and sinusoidal emissions is a snap, making FCC and ETSI compliance a breeze.

V. Wi-Fi MCU

The ESP8266 solves Wi-Fi development issues by allowing the machine for highlighting or the wireless fidelity channel for managing from a different function processor. It can launch immediately from an additional flash if the machine is serving and the device's only processor. Any microcontroller-based device can be used as a Wi-Fi adaptor by extending internet access via wireless facility via the UART interface or the CPU AHB bridge interface.

A wireless internet connection can also assist any micro - controller reliable device. The ESP8266EX is a well-designed and equiped with WiFi SoC which matches customer objectives for low power consumption, less weight, and equipped with consistent action in the IoT field. Because of its extensive and self-sufficient Wireless fidelity networking capabilities, the ESP8266EX can be used as an independent system or as a host MCU slave. Because the ESP8266EX throws the script, and application runs right instantly. The integrated super-speed cache improves system stability and memory utilisation. The ESP8266EX can be used as an SPI/SDIO or I2C controller as well.

Designs of micro controllers have the UART interface as a Wireless fidelity adapter. The ESP8266EX includes switches for antennas, RF, controlled amplifier, filters, amplifier, and power controlled devices. The small design reduces the total size of the

designed PCB, and just a few outer circuits are required.

The ESP8266EX has an stable version of Ten silica's 32-bit based Processor called diamond processor as well as on-chip SRAM memory, in addition to Wi-Fi capabilities. GPIOs can be used to link various outer sensors and other devices to it. Sample code for a range of works is provided through the Software Development Kit (SDK).

Using common wireless, Bluetooth, DD Radiation (DDR), LVDS, and LCD interruption management, the Descriptive Systems Smart Communication Platform (ESCP) allows for advanced data processing, spur-calling, and radio co-existence techniques to effortlessly switch back and forth to disrupted napping and raise methods.

VI. RESULTS

The RF transmitter section and the RF receiver section is displayed at the figure 5 and figure 6 respectively. Also Wi-Fi module is discussed in the above section and the hardware setup of the transmitter section and the receiver section is shown in the figure 7.

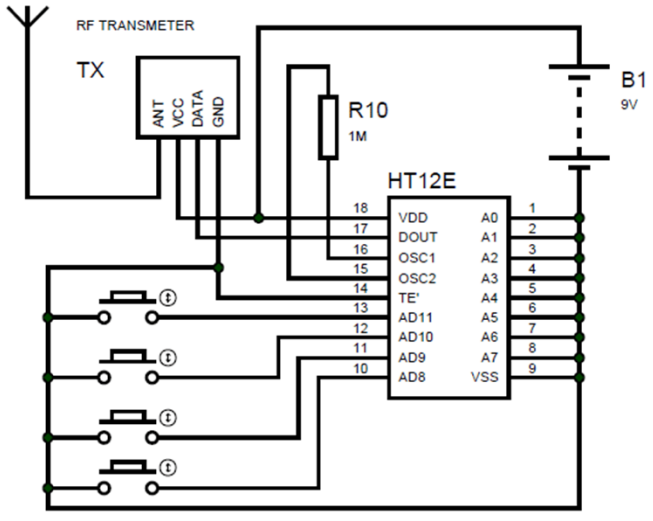


Figure 6 RF Transmitter section

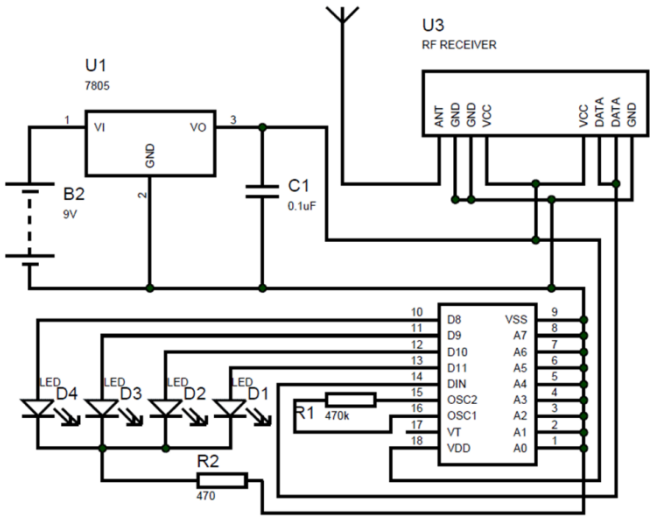


Figure 7 RF Receiver section

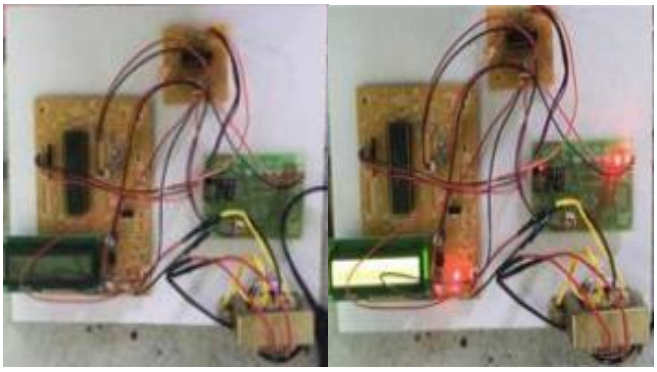


Figure 8 Hardware Setup

VII. CONCLUSIONS

The results were based on weather circumstances as well as expected scenarios such as accidents or traffic jams, and the results were based on local road monitoring with warning messages to improve traffic. It was created by combining all of the hardware components that were used. Each module's existence was meticulously designed and precisely positioned to ensure the device's best performance. Furthermore, cutting-edge ICs and technology were used to finish the work efficiently.

VIII. REFERENCES

[1]. Thakur TT, Naik A, Vatari S, Gogate M. Real time traffic management using Internet of Things. In 2016 International Conference on

Communication and Signal Processing (ICCSP) 2016 Apr 6 (pp. 1950-1953). IEEE.

[2]. Sugumar K, Kumar AS. IoT Concept for Animal Detection Using ANN to Prevent Animal Vehicle Collision on Highways. International Journal of Science and Engineering Applications Volume 8–Issue 11,492-497, 2019.

[3]. Saad W, Alsayyari AS. Real Time IoT Based Camel-Vehicle Collision Avoidance System for KSA. In 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS) 2020 Mar 19 (pp. 1-4). IEEE.

[4]. Clevenger AP, Chruszcz B, Gunson KE. Highway mitigation fencing reduces wildlife- vehicle collisions. Wildlife Society Bulletin. 2001 Jul 1:646-53.

[5]. Misbahuddin S, Zubairi JA, Saggaf A, Basuni J, Sulaiman A, Al-Sofi A. IoT based dynamic road traffic management for smart cities. In 2015 12th International conference on high- capacity optical networks and enabling/emerging technologies (HONET) 2015 Dec 21 (pp. 1-5). IEEE.

[6]. Masum AK, Chy MK, Rahman I, Uddin MN, Azam KI. An Internet of Things (IoT) based smart traffic management system: a context of Bangladesh. In 2018 International Conference on Innovations in Science, Engineering and Technology (ICISET) 2018 Oct 27 (pp. 418-422). IEEE.

[7]. Roy A, Siddiquee J, Datta A, Poddar P, Ganguly G, Bhattacharjee A. Smart traffic & parking management using IoT. In 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) 2016 Oct 13 (pp. 1-3). IEEE.

[8]. Pavithira, Deepa R. Iot in Healthcare, International Research Journal of Engineering and Technology. 2017, 4(10), pp: 931- 933.

[9]. Elkin D, Vyatkin V. IoT in Traffic Management: Review of Existing Methods of Road Traffic Regulation. In Computer Science On-line Conference 2020 Jul 15 (pp. 536-551). Springer, Cham.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Enhanced Password Processing Scheme Based On Visual Cryptography and OCR

Dr. P. Mangayarkarasi[1], Rakesh H P [2]

[1]Associate Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India
[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Customary secret phrase change conspires for client verification is to change the passwords into hash esteems. These hash-based secret phrase plans are similarly straightforward and quick on the grounds that those depend on text and celebrated cryptography. Notwithstanding, those can be presented to digital assaults using secret word by breaking instrument or hash- breaking on the web locales. Aggressors can altogether sort out a unique secret word from hash esteem when that is generally straightforward and plain. Thus, many hacking mishaps have been happened prevalently in frameworks embracing those hash- based plans. In this work, we recommend upgraded secret key preparing plan dependent on picture utilizing visual cryptogra- phy (VC). Not quite the same as the customary plan dependent on hash and text, our plan changes a client ID of text type to two pictures scrambled by VC. The client should make two pictures comprised of subpixels by irregular capacity with SEED which incorporates individual data. The worker just has client's ID and one of the pictures rather than secret word. At the point when the client signs in and sends another picture, the worker can extricate ID by using OCR (Optical Character Recognition). Therefore, it can validate client by contrasting removed ID and the saved one. Our proposition has lower calculation, forestalls digital assault pointed toward hash cracking, and supports confirmation not to uncover individual data like ID to assailants.

I. INTRODUCTION

Client validation overall frameworks has continued essen- tially through check of the ID and secret key. To send and check secret word, the framework utilizes a hash-based secret key plan that changes unique secret word to hash esteem by acclaimed work. The benefits are that it can be adapted in framework without trouble, and computational speed of cycle is quick on the grounds that a sort of hash-put together plan is generally based with respect to message using famous hash capacity like MD5, SHA256. Be that as it may, it is helpless against assaults, for example, beast power assault or word reference based assault clearly by secret key breaking device or hash-breaking on the web destinations. Accept that somebody characterizes secret word "1qaz2wsx" in a framework. On the off chance that an assailant knows about the hash esteem " 1c63129ae9db9c60c3e8aa94d3e00495", the worth can be adequately broken essentially by free break site . Despite the fact that the aggressor doesn't have a clue about any data about hash capacity, the individual can without much of a stretch supposition which sort

of hash work is adjusted in the framework. As the outcome, the aggressor can make optional harm the framework.

Problem Statement

Despite the fact that the aggressor doesn't have the foggiest idea about any data about hash capacity, the person can without much of a stretch theory which sort of hash work is adjusted in the framework. As the outcome, the assailant can make auxiliary harm the framework. Members have the obligation on this sort of assaults. At the point when an analyst asked to numerous individuals about secret word the executives practices.

II. SYSTEM ANALYSIS

INTRODUCTION TO SYSTEM ANALYSIS

System

A framework is an efficient gathering of associated parts connected together as indicated by an arrangement to accom- plish a particular goal. Its principle qualities are association, communication, reliance, coordination and a focal target.

System Analysis

Framework examination and configuration are the utilization of the framework way to deal with critical thinking for the most part utilizing PCs. To recreate a framework the expert should consider its components yield and data sources, processors, controls, criticism and climate.

Analysis

Examination is an itemized investigation of the different activities performed by a framework and their connections inside and outside of the framework. One part of examination is characterizing the limits of the framework and deciding if an up-and-comer framework ought to think about other related frameworks. During investigation information are gathered on the accessible documents choice focuses and exchanges took care of by the current framework. This includes gathering data and utilizing organized instruments for investigation.

III. EXISTING SYSTEM

Examination is an itemized investigation of the different activities performed by a framework and their connections inside and outside of the framework. One part of examination is characterizing the limits of the framework and deciding if an up-and-comer framework ought to think about other related frameworks. During investigation information are gathered on the accessible documents choice focuses and exchanges took care of by the current framework. This includes gathering data and utilizing organized instruments for investigation.

IV. PROPOSED SYSTEM

1) The user inputs the ID and password.
2) The device of user creates an original image composed of black characters and white background. If the saved original image exists on user's device, it dose not have to create the original image again.
3) Although the device does not possess the first shared im- age, it can thoroughly construct second shared image referred to the original image and first shared image because the device already knows the SEED to make up the first shared image.
4) The user sends the second shared image only to the server.
5) The server overlaps the first shared image saved and the second shared image received.
6) The server should remove the background of the over- lapped image as in Figure 3 (d), to gain original image.

7) ID is retrieved from the background-removed image by OCR .
8) The server confirms whether the extracted ID corresponds with saved ID, and determines success or fail.
9) The result is sent to the user.

Advantages of the Proposed System

The goal of our proposal is to prevent cyber attack and support privacy of personal information.

Project Module Description Modules

Visual Cryptography Technique:

In this module showed a visual secret sharing scheme, where a picture was separated into n shares so just somebody with all n shares could unscramble the picture, while any n– 1 shares uncovered no data about the first picture. Each offer was imprinted on a different straightforwardness, and decoding was performed by overlaying the offers. When all n shares were overlaid, the first picture would show up. There are a few speculations of the fundamental plan including k-out-of-n visual cryptography.

Region Growing Algorithm:

It is likewise delegated a pixel-based picture division strat- egy since it includes the choice of initial seed focuses. This way to deal with division analyzes adjoining pixels of intro- ductory seed focuses and decides if the pixel neighbors ought to be added to the area. The cycle is iterated on, in a similar way as general data clustering algorithms.

OCR (optical character recognition)

OCR (optical character acknowledgment) is the acknowl- edgment of printed or composed content characters by a PC. This includes photograph checking of the content character-by- character, examination of the filtered in picture, and afterward interpretation of the character picture into character codes, like ASCII, regularly utilized in data processing.

V. SYSTEM REQUIREMENTS

Functional Requirements

This System must be created in MVC Architecture on java innovation. There ought to be two entertainers Admin, User.

- Admin ought to have separate login meeting where she/he can ready to login with Username and secret word and ready to keep up User Creation subtleties.
- At the hour of client creation itself client accepting offer picture to client register email and offer (II) administrator putting away in Server.
- User needs to login by utilizing username and he needs to get the verify picture in worker application as offer (II).
- Already client got share (I) picture from their email and offer (II) worker App now following stage he needs to blend the share(I) and share(II) picture then, at that point removing the userid utilizing OCR method we are doing userid confirmation measure.

Non Functional Requirements

- **Usability**

Basic is the key here. The framework should be basic that individuals like to utilize it, however not so intricate that individuals try not to utilize it. The client should be acquainted with the UIs and ought not have issues in relocating to another framework with another climate. The menus, catches and exchange boxes ought to be named in a way that they give clear comprehension of the usefulness. A few clients will utilize the framework all the while, so the convenience of the framework ought not get influenced concerning singular clients.

- **Reliability**

The framework ought to be dependable and solid in giving the functionalities. When a client has rolled out certain improvements, the progressions should be made apparent by the framework. The progressions made by the Programmer ought to be noticeable both to the Project chief just as the Test engineer.

- **Performance**

The framework will be utilized by numerous workers all the while. Since the framework will be facilitated on a solitary web worker with a solitary data set worker behind the scenes, execution turns into a significant concern. The framework ought not surrender when numerous clients would utilize it at the same time. It ought to permit quick availability to the entirety of its clients. For instance, if two test engineers are at the same time attempting to report the presence of a bug, then, at that point there ought not be any irregularity at the same time.

- **Scalability**

The framework ought to be adequately adaptable to add new functionalities at a later stage. There ought to be a typical channel, which can oblige the new functionalities.

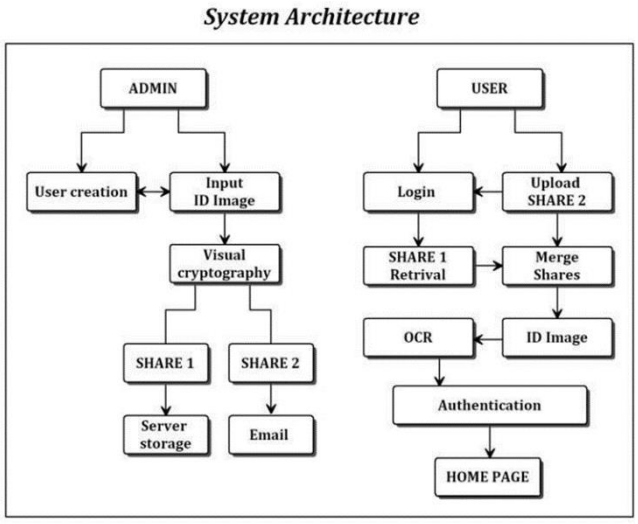
- **Maintainability**

The framework observing and support ought to be basic and objective in its methodology. There ought not be an excessive number of occupations running on various machines to such an extent that it gets hard to screen whether the positions are running without blunders.

- **Portability**

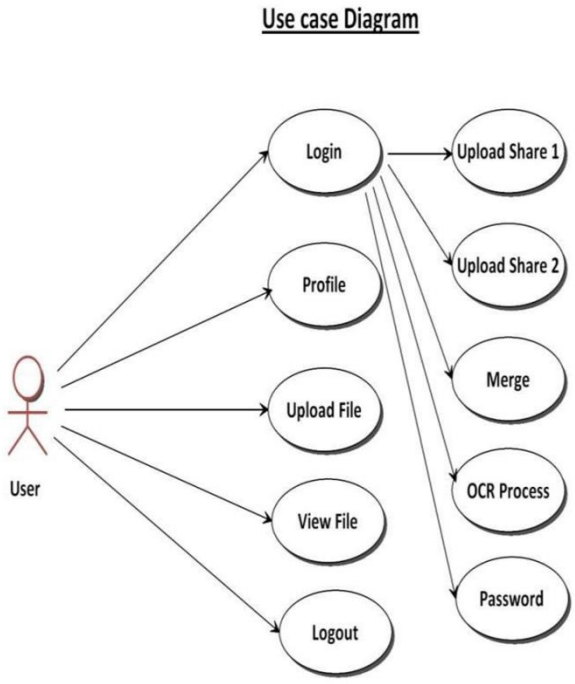
The framework ought to be effectively compact to another framework. This is required when the web worker, which s facilitating the framework stalls out because of certain issues, which requires the framework to be taken to another framework.

VI. SYSTEM ARCHITECTURE

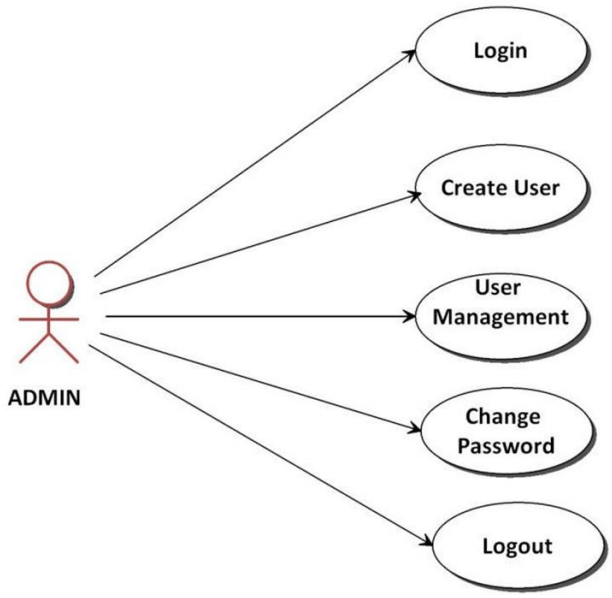


USE CASE DIAGRAM

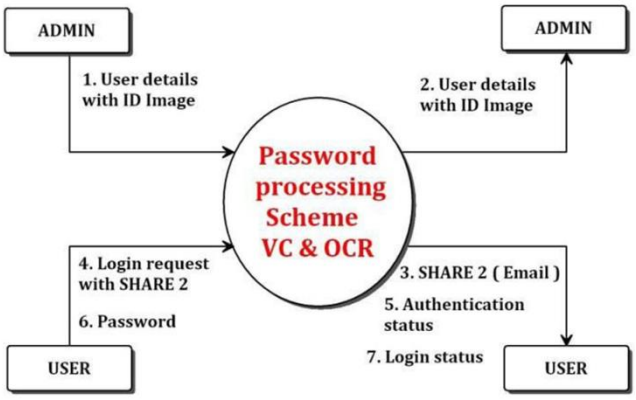
Admin:



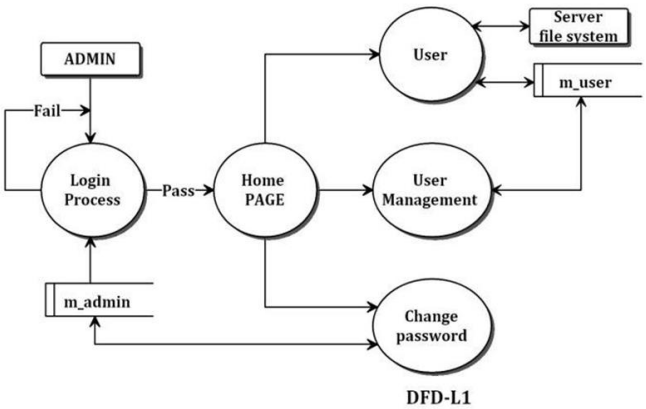
Use case Diagram



Contaxt Analysis Diagram



DFD - Admin Session



DFD-L1

IMPLEMENTATION TECHNOLOGIES

The execution stage includes something beyond composing code. Code additionally should be tried and fixed just as arranged and incorporated into a total executable item. We normally need to use design the executives to monitor diverse form of code. This is the phase of the undertaking where the hypothetical plan is transformed into a functioning framework. On the off chance that the execution isn't painstakingly arranged and controlled, it can cause bedlam and disarrays. It is consistently a smart thought to remember that a few qualities that ought to be found in a decent execution like Readability- our code is written in MVC Architecture ,JAVA to accomplish the target of the undertaking that is to present a novel plan of component plan for adjusting the asset utilizations .

Our execution stage requires the accompanying undertakings:

- Careful arranging
- Investigation of framework and limitations
- Design of techniques to accomplish the changeover
- Evaluation of the changeover technique
- Correct choices in regards to determination of the stage
- Appropriate determination of the language for application improvement Java Technology Java innovation is both a programming language and a stage.

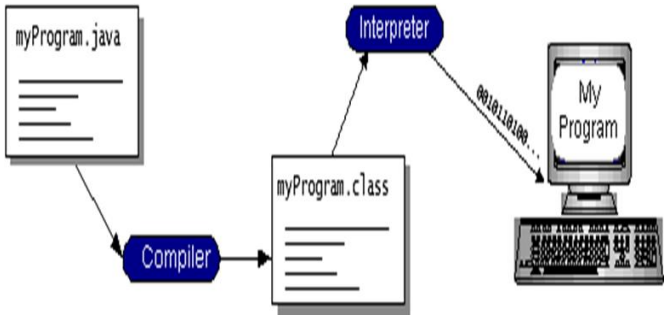
THE JAVA PROGRAMMING LANGUAGE

The Java programming language is a significant level language that can be portrayed by the entirety of the accompanying popular expressions:

- Simple
- Architecture nonpartisan
- Object arranged
- Portable

- Distributed
- High execution
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming dialects, you either accumulate or decipher a program so you can run it on your PC. The Java programming language is surprising in that a program is both accumulated and deciphered. With the compiler, first you make an interpretation of a program into a transitional language called Java byte codes — the stage free codes deciphered by the translator on the Java stage. The translator parses and runs every Java byte code guidance on the PC. Aggregation happens only a single time; understanding happens each time the program is executed. The accompanying figure outlines how this functions.



You can consider Java byte codes as the machine code directions for the Java Virtual Machine (Java VM). Each Java mediator, regardless of whether it's an improvement apparatus or a Web program that can run applets, is an execution of the Java VM. Java byte codes help make "compose once, run anyplace" conceivable. You can assemble your program into byte codes on any stage that has a Java compiler. The byte codes would then be able to be run on any execution of the Java VM. That implies that up to a PC has a Java VM, a similar program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

VII. REFERENCES

[1]. Brian Cluster analysis. , West Sussex, U.: Wiley. ISBN 9780470749913, 2011.

[2]. Silva, , "Practical Eclipse Rich Client Platform Projects (1st ed.)". . p. 352. ISBN 1-4302-1827-4, March 2009.

[3]. Riedl, C.; Zanibbi, R.; Hearst, M. A.; Zhu, S.; Menietti, M.; Crusan, J.; Metelsky, I.; Lakhani, K. (February 20, 2016). "Detecting Figures and Part Labels in Patents: Competition-Based Development of Image Processing Algorithms". International Journal on Document Analysis andRecognition. 19 (2):155.

[4]. Gaw, Shirley, and Edward W. Felten, "Password management strategies for online accounts," Proceedings of the second symposiumon Usable privacy and security. ACM, 2006.

[5]. Nguyen, Thi Thu Trang, and Quang Uy Nguyen, "An analysis of Persuasive Text Passwords, "Information and Computer Science (NICS), 2015 2nd National Foundation for Science and Technology Development Conference on. IEEE,2015.

[6]. Tam, Leona, Myron Glassman, and Mark Vandenwauver, "The psychology of password management: a tradeoff between security and convenience, "Behaviour & Information Technology 29.3 (2010): 233- 244.

[7]. Wang, Luren, Yue Li, and Kun Sun, "Amnesia: A Bilateral Generative Password Manager," 2016 IEEE 36th International Conference on Distributed Computing Systems.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

An Efficient and Privacy-Preserving Biometric Identification Scheme in Cloud Computing

Dr. K. Saravanan[1], Punith M [2]
[1]Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India
[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Biometric recognizable proof has gotten progressively well-known as of late. With the turn of events of distributed computing, data set proprietors are propelled to revaluate the huge size of biometric information and ID undertakings to the cloud to dispose of the costly stockpiling and calculation costs, which, be that as it may, carries possible dangers to clients' security. In this paper, we propose an effective and security saving bio-metric ID rethinking plan. In particular, the biometric to execute a biometric ID, the information base proprietor scrambles the inquiry information also, submits it to the cloud. The cloud performs recognizable proof activities over the scrambled data set and returns the outcome to the data set proprietor. A careful security investigation demonstrates that the proposed conspire is secure regardless of whether assailants can fashion ID demands and conspire with the cloud.

I. INTRODUCTION

Biometric recognizable proof has raised progressively confirm the character of a person. Data from numerous sources consideration since it gives a promising method to can be combined in a few particular levels, including the distinguish clients. Contrasted and conventional component extraction level, match score level and choice level. confirmation techniques dependent on passwords and While combination at the match score and choice levels have distinguishing proof cards, biometric ID is viewed as more been broadly concentrated in the writing, combination at the solid and helpful. Furthermore, biometric distinguishing component level is a generally understudied issue. In this paper proof has been generally applied in numerous fields by we talk about combination at the component level in 3 unique utilizing biometric attributes like finger impression, iris and facial examples, which can be gathered from different sensors .In a biometric ID framework, the data set proprietor, for example, the FBI who is mindful to deal with the public fingerprints data set, may want to revaluate the colossal biometric information to the cloud worker (e.g., Amazon) to dispose of the costly stockpiling and calculation costs. Nonetheless, to save the security of biometric information, the biometric information must be scrambled prior to re-appropriating. At whatever point a FBI's accomplice (e.g., the police headquarters) needs to verify a person's character, he goes to the FBI and creates an ID inquiry by utilizing the person's biometric attributes (e.g., fingerprints, irises, voice designs,

facial examples and so Less Security: During the recognizable proof interaction, the on) Then, at that point, the FBI scrambles the question and security of biometric information ought not be ensured. submits it to the cloud to track down the nearby match. Assailants and the semi-genuine cloud ought to become familiar Along these lines, the difficult issue is the way to plan a with about the touchy data, there is no security to ensure our convention which empowers productive and protection information. Saving biometric distinguishing proof in the distributed computing. Various protection safeguarding biometric recognizable proof arrangements have been proposed.

Problem Statement

In July 2018 telecom administrative authority of India (trail) director R.S Sharma post is creator no in twitter and difficulties creator pundits to do him hurt on the off chance that they could. Inside 7 hours moral programmers posted screen capture of sending re.1 to Sharma through the adhar empowered help and furthermore they distributed 14 things, which incorporates Sharma's portable no DOB, private location, telephone no, PAN no, Bank subtleties, and so on As of now our Aadhaar card information.

II. SYSTEM ANALYSIS

INTRODUCTION TO SYSTEM ANALYSIS

Multi biometric frameworks use the proof introduced by different biometric sources (e.g., face and finger impression, various fingers of a client, numerous matchers, and so forth) to decide or situations:

(i) combination of PCA and LDA coefficients of face;

(ii) combination of LDA coefficients relating to the R,G,B channels of a face picture;

(iii) Combination of face and hand modalities. Starter results are empowering and help in featuring the upsides and downsides of performing combination at this level. The essential inspiration of this work is to exhibit the practicality of such a combination and to highlight the significance of seeking after additional examination toward this path.

III. PROPOSED SYSTEM

We propose a proficient and protection safeguarding biometric ID conspire which can oppose the conspiracy assault dispatched by the clients and the cloud. We analyze the biometric ID plan and show it's in adequacy's and security shortcoming under the proposed level-3 assault. In particular, we exhibit that the aggressor can recuperate their mysterious keys by intriguing with the cloud, and afterward decode the biometric qualities of all clients three kinds of elements are engaged with the framework including the data set proprietor, clients and the cloud. The data set proprietor holds an enormous size of biometric information which is scrambled and sent to the cloud for capacity. At the point when a client needs to recognize him/her, a question demand is be shipped off the data set proprietor. In the wake of getting the solicitation, the data set proprietor creates a ciphertext for the biometric characteristic and afterward communicates the ciphertext to the cloud for distinguishing proof. The cloud worker sorts out the best counterpart for the scrambled question and returns the connected file to the data set proprietor. At last, the data set proprietor figures the comparability between the inquiry information and the biometric information related with the record, and returns the question result to the client.

Advantages of the Proposed System

- More security with Block chain stockpiling
- Reduce responsibility and upgrade usefulness
- Better adaptability and speed

- Efficiency: Computational expenses ought to be pretty much as low as conceivable at both the data set proprietor side and the client side. To acquire high effectiveness, most biometric distinguishing proof tasks ought to be executed in the cloud.
- Security: During the recognizable proof interaction, the security of biometric information ought to be ensured. Aggressors and the semi-legitimate cloud ought to adapt nothing about the delicate data.

IV. SYSTEM REQUIREMENTS

A Software Requirement Specification (SRS) is essentially an association's comprehension of a client or potential customer's framework necessities and conditions at a specific point preceding any genuine plan or improvement work. The data accumulated during the investigation is converted into a report that characterizes arrangements of prerequisites. It gives the short depiction of the administrations that the framework ought to give and furthermore the imperatives under which, the framework ought to work. For the most part, the SRS is a report that totally depicts what the proposed programming ought to manage without portraying how the product will do it. It's a two-way protection strategy that guarantees that both the customer and the association comprehend different are necessities from that viewpoint at a given point on schedule.

The SRS report itself states in exact and unequivocal language those capacities and abilities a product framework should give, just as states any necessary imperatives by which the framework should withstand. The SRS additionally works as an outline for finishing a venture with as little expense development as could be expected. The SRS is frequently alluded to as the "parent" archive since all resulting project the board records, for example, plan particulars, explanations of work, programming engineering determinations, testing and approval plans, and documentation plans, are identified with it. Necessity is a condition or ability to which the framework should adjust. Prerequisite Management is an orderly methodology towards inspiring, putting together and recording the necessities of the framework plainly alongside the appropriate properties. The tricky troubles of Requirements are not generally self- evident and can emerge out of quite a few sources.

Non Functional Requirements

- **Usability**

Basic is the key here. The framework should be basic that individuals like to utilize it, however not so intricate that individuals try not to utilize it. The client should be acquainted with the UIs and ought not have issues in relocating to another framework with another climate. The menus, catches and exchange boxes ought to be named in a way that they give clear comprehension of the usefulness. A few clients will utilize the framework all the while, so the convenience of the framework ought not get influenced concerning singular clients.

- **Reliability**

The framework ought to be dependable and solid in giving the functionalities. When a client has rolled out certain improvements, the progressions should be made apparent by the framework. The progressions made by the Programmer ought to be noticeable both to the Project chief just as the Test engineer.

- **Performance**

The framework will be utilized by numerous workers all the while. Since the framework will be facilitated on a solitary web worker with a solitary data set worker behind the scenes, execution turns into a significant concern. The framework ought not surrender when numerous clients would utilize it at

the same time. It ought to permit quick availability to the entirety of its clients. For instance, if two test engineers are at the same time attempting to report the presence of a bug, then, at that point there ought not to be any irregularity at the same time.

- **Scalability**

The framework ought to be adequately adaptable to add new functionalities at a later stage. There ought to be a typical channel, which can oblige the new functionalities.

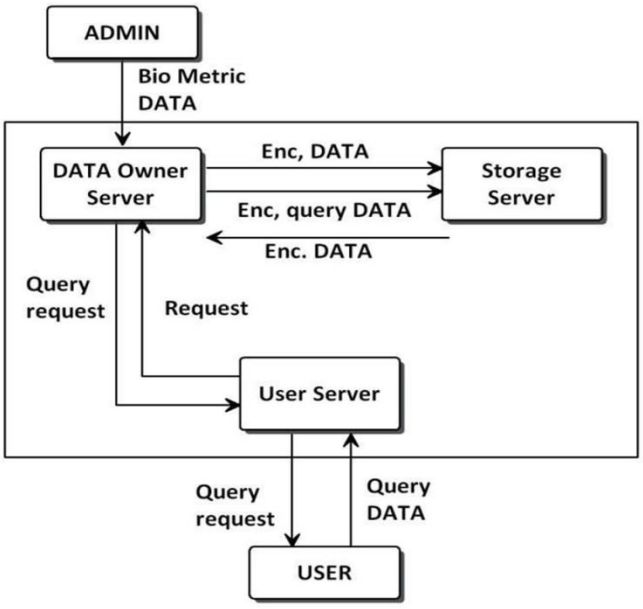
- **Maintainability**

The framework observing and support ought to be basic and objective in its methodology. There ought not to be an excessive number of occupations running on various machines to such an extent that it gets hard to screen whether the positions are running without blunders.

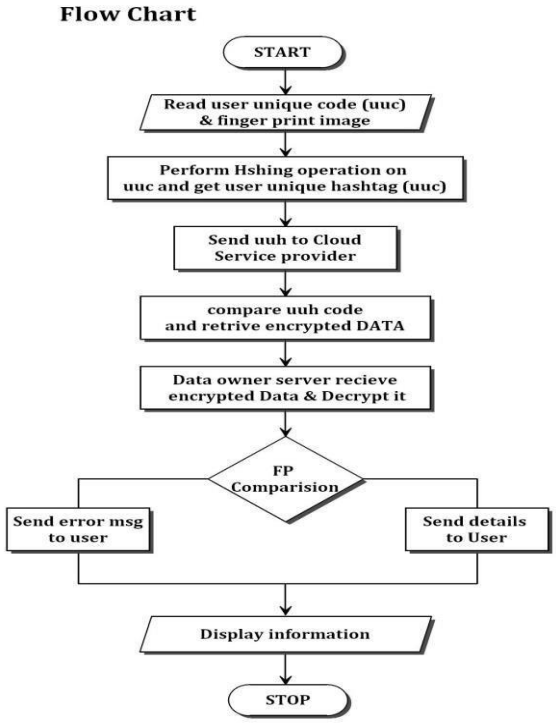
- **Portability**

The framework ought to be effectively compact to another framework. This is required when the web worker, which s facilitating the framework stalls out because of certain issues, which requires the framework to be taken to another framework.

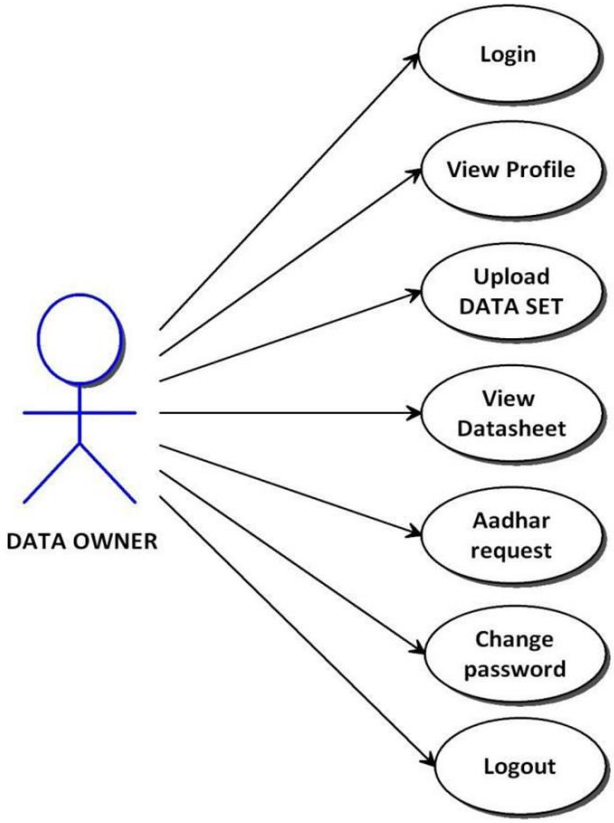
V. SYSTEM ARCHITECTURE

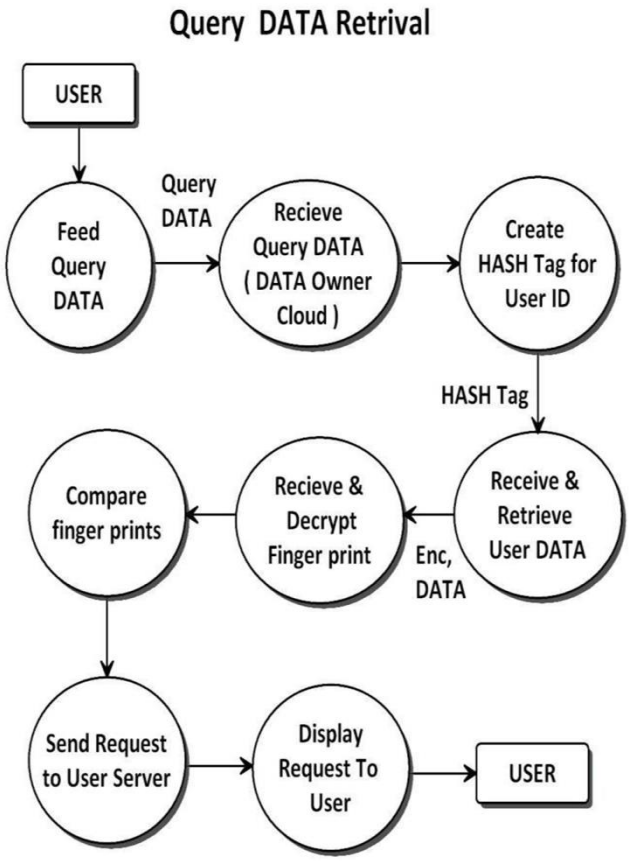


FLOWCHART DIAGRAM



USE CASE DIAGRAM





IMPLEMENTATION TECHNOLOGIES

The execution stage includes something other than composing code. Code additionally should be tried and repaired just as ordered and incorporated into a total executable item. We as a rule need to use arrangement the executives to monitor diverse form of code. This is the phase of the venture where the hypothetical plan is transformed into a functioning framework. In the event that the execution isn't painstakingly arranged and controlled, it can cause bedlam and disarrays. It is consistently a smart thought to remember that a few attributes that ought to be found in a decent execution like Readability- our code is written in MVC Architecture, JAVA to accomplish the goal of the undertaking that is to present a novel plan of system plan for adjusting the asset utilizations.

Our execution stage requires the accompanying errands:

- Careful arranging
- Investigation of framework and requirements
- Design of strategies to accomplish the changeover
- Evaluation of the changeover strategy
- Correct choices with respect to determination of the stage
- Appropriate choice of the language for application improvement Java Technology Java innovation is both a programming language and a stage.

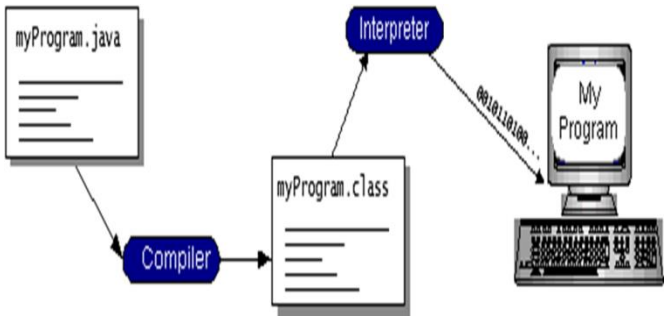
THE JAVA PROGRAMMING LANGUAGE

The Java programming language is a significant level language that can be portrayed by the entirety of the accompanying popular expressions:

- Simple
- Architecture nonpartisan
- Object arranged
- Portable
- Distributed

- High execution
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming dialects, you either accumulate or decipher a program so you can run it on your PC. The Java programming language is surprising in that a program is both accumulated and deciphered. With the compiler, first you make an interpretation of a program into a transitional language called Java byte codes — the stage free codes deciphered by the translator on the Java stage. The translator parses and runs every Java byte code guidance on the PC. Aggregation happens only a single time; understanding happens each time the program is executed. The accompanying figure outlines how this functions.



You can consider Java byte codes as the machine code directions for the Java Virtual Machine (Java VM). Each Java mediator, regardless of whether it's an improvement apparatus or a Web program that can run applets, is an execution of the Java VM. Java byte codes help make "compose once, run anyplace" conceivable. You can assemble your program into byte codes on any stage that has a Java compiler.

VI. REFERENCES

[1]. A. Jain, L. Hong, and S. Pankanti, "Biometric identification," Commun. ACM, vol. 43, no. 2, pp. 90–98, 2000.

[2]. R. Allen, P. Sankar, and S. Prabhakar, "Fingerprint identification technology," in Biometric Systems. London, U.K.: Springer, 2005, pp. 22–61.

[3]. J. de Mira, Jr., H. V. Neto, E. B. Neves, and F. K. Schneider, "Biometric oriented iris identification based on mathematical morphology," J. Signal Process. Syst., vol. 80, no. 2, pp. 181–195, 2015.

[4]. S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3D morphable model using linear shape and texture error functions," in Proc. Eur. Conf. Comput. Vis., 2002, pp. 3–19.

[5]. Y. Xiao et al., "A survey of key management schemes in wireless sensor networks," Comput. Commun., vol. 30, nos. 11–12, pp. 2314–2341, Sep. 2007.

[6]. X. Du, Y. Xiao, M. Guizani, and H.-H. Chen, "An effective key management scheme for heterogeneous sensor networks," Ad Hoc Netw., vol. 5, no. 1, pp. 24–34, 2007.

[7]. X. Du and H. H. Chen, "Security in wireless sensor networks," IEEE Wireless Commun. Mag., vol. 15, no. 4, pp. 60–66, Aug. 2008.

[8]. X. Hei and X. Du, "Biometric-based two-level secure access control for implantable medical devices during emergencies," in Proc. IEEE INFOCOM, Apr. 2011, pp. 346–350.

[9]. X. Hei, X. Du, J. Wu, and F. Hu, "Defending resource depletion attacks on implantable medical devices," in Proc. IEEE GLOBECOM, Dec. 2010, pp. 1–5.

[10]. M. Barni et al., "Privacy-preserving fingercode authentication," in Proc. 12th ACM Workshop Multimedia Secur., 2010, pp. 231–240.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Security Challenges in Data Analytics

Akansha

PG Scholar, Cyber Forensics and Information Security and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

The measure of information in world is developing step by step. Information is developing a result of utilization of web, brilliant telephone and interpersonal organization. Huge information is an assortment of informational collections which is enormous in size just as intricate. By and large size of the information is Petabyte and Exabyte. Conventional data set frameworks can't catch, store and break down this huge measure of information. As the web is developing, measure of huge information keep on developing. Huge information examination give new ways for organizations and government to break down unstructured information. Presently a days, Data information is perhaps the most talked subject in IT industry. It will assume significant part in future. Enormous information changes how information is overseen and utilized. A portion of the applications are in regions like medical services, traffic the board, banking, retail, schooling, etc. Associations are getting more adaptable and more open. New sorts of information will give new difficulties too. The present paper features significant ideas of Data Analytics.

I. INTRODUCTION

The term Data Analytics is currently utilized wherever in our day by day life. The term Data Analytics came around 2005 which alludes to a wide scope of enormous informational indexes practically difficult to oversee and deal with utilizing customary information the board devices – due to their size, yet additionally their intricacy. Huge Data can be found in the money and business where huge measure of stock trade, banking, on the web and on location buying information courses through modernized frameworks consistently and are then, at that point caught and put away for stock checking, client conduct and market conduct. It can likewise be found in the life sciences where large arrangements of information, for example, genome sequencing, clinical information and patient information are broke down and used to advance forward leaps in science in The Data Analytics scene can be partitioned into two primary research. Different spaces of examination where Data classes: Systems which give operational abilities to continuous, Analytics is of focal significance are cosmology, conditional/intuitive circumstances where information is caught oceanography, and designing among numerous others. The and put away. The other kind is frameworks that give jump in computational and capacity power empowers the investigation abilities to review and complex examination of the assortment, stockpiling and examination of these Big Data information that has been put away. This archive is a layout. An sets and organizations acquainting scientists are zeroing in on comparable to it.

Problem Statement

The Data Analytics scene can be partitioned into two fundamental classifications: Systems which give operational abilities to continuous, value-based/intuitive circumstances where information is caught and put away. The other sort is frameworks that give examination capacities to review and complex investigation of the information that has been put away. This archive is a layout. An electronic duplicate can be downloaded from the Journal site. For inquiries on paper rules, kindly contact the diary distributions council as demonstrated on the diary site. Data about conclusive paper accommodation is accessible from the meeting site. The accompanying table is a correlation among Operation and Logical Systems in the field of Data Analytics.

II. SYSTEM ANALYSIS

INTRODUCTION TO SYSTEM ANALYSIS

Enormous information investigation alludes to the way toward gathering, coordinating and breaking down huge arrangements of information ("huge information") to find designs and other helpful data. With the assistance of Data Analytics investigation, associations utilize the huge measures of information made accessible to them to distinguish examples and concentrate valuable data. Enormous Data examination not just assists us with understanding the data contained in the information yet additionally recognize the data that is most imperative to the association and future choices. The main objective of Data Analytics is to empower associations to settle on better choices. Information Scientists, prescient modelers and other investigation experts bargain Inventive mechanical electronic duplicate can be downloaded from the Journal site. For arrangements with Huge Data investigation are prospering. Inquiries on paper rules, if it's not too much trouble, contact the scientists are zeroing in on comparable to it.

III. EXISTING SYSTEM

In this article, we investigate the term Data Analytics as it diary distributions advisory group as shown on the diary site. rose up out of the friend surveyed writing. Rather than news Data about conclusive paper accommodation is accessible from things and web-based media articles, peer inspected articles the meeting site. The accompanying table is an examination offer a brief look into Data Analytics as a subject of study among Operation and Logical Systems in the field of Data and the logical issues systems and arrangements that Analytics.

IV. PROPOSED SYSTEM

A definitive advance in Data Analytics preparing incorporates translation and acquiring important data from the information that is prepared. The data acquired can be of two sorts: Retrospective Investigation incorporates acquiring bits of knowledge about occasions and moves that have effectively made spot. For example, information about the TV viewership for a show in various regions can help us judge the prevalence of the show in those regions. Planned Analysis incorporates passing judgment on designs and knowing patterns for future from information that is as of now been produced. Climate Prediction utilizing large information examination is an illustration of planned investigation. Issues building from such understandings relate to erroneous and misdirecting patterns being anticipated. This is especially risky because of an expanding dependence on information for key choices. For instance, if a specific indication is plotted against the probability of being determined to have a specific illness; it may prompt deception about the manifestation being caused because of the specific infection itself. Bits of knowledge acquired from information translation are along these lines vital and the essential justification preparing huge information

as well. All sections should be indented. All passages should be supported, for example left-supported and right- advocated.

Advantages of the Proposed System

- ❖ More security with Block chain stockpiling
- ❖ Reduce responsibility and upgrade usefulness
- ❖ Better adaptability and speed
- ❖ Efficiency: Computational expenses ought to be pretty much as low as conceivable at both the data set proprietor side and the client side. To acquire high effectiveness, most biometric distinguishing proof tasks ought to be executed in the cloud.

V. SYSTEM REQUIREMENTS

A Software Requirement Specification (SRS) is basically an affiliation's understanding of a customer or potential client's structure necessities and conditions at a particular point going before any certifiable arrangement or improvement work. The information amassed during the examination is changed over into a report that portrays a game plans of essentials. It gives the short portrayal of the organizations that the system should give and moreover the goals under which, the structure should work. Generally, the SRS is a report that absolutely portrays what the proposed programming should oversee without depicting how the item will do it. It's a two-way security technique that ensures that both the client and the affiliation appreciate various' necessities from that perspective at a given point on time.

The SRS report itself states in definite and unequivocal language those limits and capacities an item system should give, similarly as states any important goals by which the structure ought to withstand. The SRS also functions as a layout for completing an endeavor with as little cost improvement as could be anticipated. The SRS is habitually implied as the "parent" file since all subsequent undertaking the board records, for instance, plan points of interest, clarifications of work, programming designing conclusions, testing and endorsement plans, and documentation plans, are related to it. Need is a condition or capacity to which the system ought to change. Essential Management is an efficient philosophy towards rousing, assembling and recording the necessities of the structure clearly close by the fitting properties. The interesting difficulties of Requirements are not for the most part undeniable and can arise out of a significant number sources.

Non Functional Requirements

- **Usability**

Basic is the key here. The framework should be basic that individuals like to utilize it, however not so intricate that individuals try not to utilize it. The client should be acquainted with the UIs and ought not have issues in relocating to another framework with another climate. The menus, catches and exchange boxes ought to be named in a way that they give clear comprehension of the usefulness. A few clients will utilize the framework all the while, so the convenience of the framework ought not get influenced concerning singular clients.

- **Reliability**

The framework ought to be dependable and solid in giving the functionalities. When a client has rolled out certain improvements, the progressions should be made apparent by the framework. The progressions made by the Programmer ought to be noticeable both to the Project chief just as the Test engineer.

- **Performance**

The framework will be utilized by numerous workers all the while. Since the framework will be facilitated on a solitary web worker with a solitary data set

worker behind the scenes, execution turns into a significant concern. The framework ought not surrender when numerous clients would utilize it at the same time. It ought to permit quick availability to the entirety of its clients. For instance, if two test engineers are at the same time attempting to report the presence of a bug, then, at that point there ought not be any irregularity at the same time.

- **Scalability**

The framework ought to be adequately adaptable to add new functionalities at a later stage. There ought to be a typical channel, which can oblige the new functionalities.

- **Maintainability**

The framework observing and support ought to be basic and objective in its methodology. There ought not be an excessive number of occupations running on various machines to such an extent that it gets hard to screen whether the positions are running without blunders.

- **Portability**

The framework ought to be effectively compact to another framework. This is required when the web worker, which s facilitating the framework stalls out because of certain issues, which requires the framework to be taken to another framework.

IMPLEMENTATION TECHNOLOGIES

It is extremely reliant upon your industry and friends how you can uphold you with security computerization. In the event that it's retail, medical care, fabricating, monetary administrations, the public area, or another industry, the assets and cycles can depend intensely upon. Retailers for example bargain in flighty manners with ransomware and phishing assaults. Computerization is compelling in preparing for battle of rehashed assaults and bogus positives, which will improve security examiners ready to explore these cases and find a drawn out arrangement. It is critical to work with an IT group and other authoritative pioneers to perceive issues that should be tended to before any merchant is considered. Computerization is on the rundown of need regions as organizations comprehend that it disposes of hazards, makes their organizations straightforward, and influences their security stacks. The decrease of human mistake is one of the best dangers. On the off chance that an architect is called upon to play out a similar undertaking each day, looking for needles in similar piles, they ultimately commit an error. Numerous business security advances and administrations are broke down to comprehend computerized controls, especially those which permit focal administration tasks to be robotized.

VI. CONCLUSION AND FUTURE SCOPE

Data is changing the manner in which we see our reality. The effect huge information has made and will keep on making can swell through all aspects of our life. Worldwide Data is on the ascent, by 2020, we would have quadrupled the information we produce consistently. This information would be produced through a wide exhibit of sensors we are consistently fusing in our lives. Information assortment would be supported by what is today named as the "Web of Things". Using keen bulbs to keen vehicles, ordinary gadgets are creating more information than any other time. These keen gadgets are joined not just with sensors to gather information surrounding them yet they are additionally associated with the network which contains different gadgets. A Smart Home today comprises of a widely inclusive engineering of gadgets that can connect with one another by means of the immense web network. Bulbs that faint consequently helped by encompassing light sensors and vehicles that can float through hefty traffic utilizing closeness sensors are

instances of sensor innovation progressions that we have seen throughout the long term. Huge Data is moreover changing things in the business world. Organizations are utilizing huge information investigation to target showcasing at quite certain socioeconomics. Center Groups are getting progressively repetitive as investigation firms, for example, McKinsey are utilizing examination on enormous example bases that have today been made conceivable because of headways in Big Data. The potential worth of worldwide individual area information is assessed to be $700 billion to end clients, and it can result in an up to half decline in item advancement and get together expenses, as indicated by a new McKinsey report. Huge Data doesn't emerge out of a vacuum: it is recorded from some information producing source. For instance, think about our capacity to detect and notice the world around us, from the pulse of an older resident, and presence of poisons noticeable all around we inhale, to the arranged square kilometer exhibit telescope, which will create up to 1 million terabytes of crude information each day. Essentially, logical analyses and reproductions can without much of a stretch produce petabytes of information today. A lot of this information is of no interest, and it very well may be sifted and compacted by significant degrees. There is monstrous degree in Big Data and an immense extension for research and Improvement.

VII. REFERENCES

[1]. A.U. Haq and T. S. Khan, “Security in automation: Smartphone might bethe greatest threat,” CFE Media, 2015.Retrieved from: https://www.controleng.com/articles/security-in-automation-smartphonemight-be-the-greatest-threat/

[2]. E.Barak, “Explaining security automation and its evolving definitions,“New York, NY: IDG Communications, Inc,2016.Retrieved from:https://www.networkworld.com/article/3121275/explaining-securityautomation-and-its-evolving-definitions.html

[3]. K. Panos, “Security Automation and Threat Information-Sharing Options,” IEEE Security & Privacy 12,2014,42-51.

[4]. M. Metheny, "Continuous monitoring through security automation, "ScienceDirect, 2017.Retrieved from: https://www.sciencedirect.com/topics/computer - science/securityautomation

[5]. P.Nguyen and A.Graham, “Enhancing Security with Automation and Orchestration, “Serious Edge,2015.Retrieved from:https://edge.siriuscom.com/security/enhancing-security-with-automationand-orchestration

[6]. R. Montesino and S. Fenz, "Automation Possibilities in Information Security Management," 2011 European Intelligence and Security Informatics Conference, Athens, 2011, pp. 259-262, DOI:10.1109/EISIC.2011.39.

[7]. T. AlSadhan and J. S. Park, "Enhancing Risk-Based Decisions by Leveraging Cyber Security Automation," 2016 European Intelligence and Security Informatics Conference (EISIC), Uppsala, 2016, pp. 164-167, DOI:10.1109/EISIC.2016.042.

[8]. C. N. N. Hlyne, P. Zavarsky, and S. Butakov, "SCAP benchmark for Cisco router security configuration compliance," 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), London, 2015, pp. 270-276, DOI: 10.1109/ICITST.2015.7412104.

[9]. G. B. Peterside, P. Zavarsky, and S. Butakov, "Automated security configuration checklist for a cisco IPsec VPN router using SCAP 1.2," 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), London, 2015, pp. 355-360, DOI:10.1109/ICITST.2015.7412120.

[10]. M. Brunner, C. Sillaber and R. Breu, "Towards Automation in Information Security Management Systems," 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS), Prague, 2017, pp.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Financial Markets Prediction Using Data Mining Techniques With R

Dr. Kalaivani D[1], Ganesh K[2]
[1]Associate Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India
[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

The Stock market is the place where segments of uninhibitedly recorded associations are exchanged. The offers are bought and sold depending up accessible records. The expense of stocks and assets are a huge bit of the economy. There are various parts that impact offer expenses. In any case there is no specific explanation at the expenses to rise or fall. This makes adventure subject to various risks. The expenses of things to come stocks are affected by the past and current market records. Accordingly budgetary trade desire procedures like ARIMA and ARMA are used for transient envisioning. This paper proposes a protections trade desire model subject to the examination of past data and ARIMA model. This model will assist budgetary pros with buying or sell stocks at the helpful time. The guess results are envisioned using R programming language.

Keywords : Stock Market, Data Mining, Prediction, ARIMA, Time Series Data, R

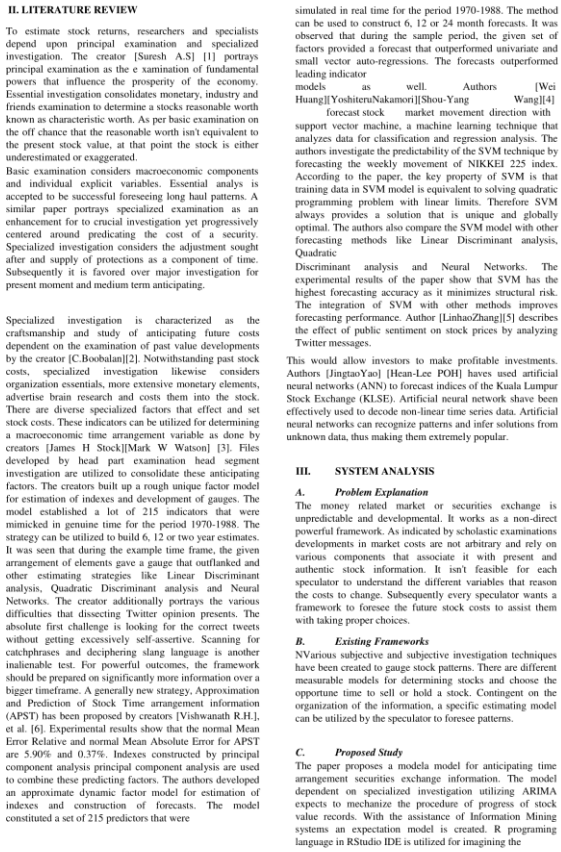
I. INTRODUCTION

The Financial market related trade structure contains 2 segments, the basic market and the discretionary market. The basic market is the place straightforwardly recorded associations offer their proposals in a first offer of stock (IPO) to raise advantages for meet their essentials of hypothesis. The helper market suggests the market where stocks are traded after their underlying contribution to individuals as a rule or in the wake of being recorded on the Stock Exchange. It is free arrangement of money related trades, not bound to any physical office or component. The expenses of the stocks depend upon market designs, adventure methodologies and other passing inefficient perspectives. This haphazardness makes it difficult to show a structure to measure stock expenses with precision. The basic doubt made while foreseeing stock data is that future market designs are affected by the stock information available unreservedly already. This suggests, the recorded stock data gives information into its future direct. As demonstrated by the Random Walk speculation for protections trades, "financial exchange costs advance as indicated by an arbitrary walk and hence can't be anticipated". The hypothesis is additionally partitioned into 2 separate parts.

The essential hypothesis communicates that reformist worth changes in an individual security are free. The ensuing hypothesis communicates the expenses conform to a particular probability transport. In any case, it is the probability flow of data or the kind of allotment that empowers academicians and examiners to appraise stock data. Late examinations have shown that Time Series data assessment

procedures give evident information to measuring stock expenses. Time plan data is progression of data accumulated over decided time span. Time game plan data for money related trade estimate can be accumulated on a step by step, after quite a while after week, month to month or yearly reason. The assessment of the time course of action data removes accommodating authentic information to grasp ascribes of data. Time game plan guaging strategies incorporate using models to anticipate future characteristics reliant on past information. R is an open source programming language and programming condition for quantifiable figuring and representations. It has different applications in the field of data assessment and for the most part used by experts and data excavators. Close by a request line interface, it has a couple of practical front-closes. R is extensible through limits, expansions and packs, contributed by the overall R society. Beginning at 2016, 7801 additional groups are open for foundation. This customer made packs like check, subtleties, ggplot2 empowers the customer to perform explicit real and graphical strategies. RStudio is an open source composed headway. Condition (IDE) for R. The item is written in C++ programming and uses Qt structure for graphical UI. It bolsters direct code execution similarly as mechanical assemblies for real examination, investigating and workspace the chiefs. There are 2 arrivals of RStudio, RStudio Desktop and RStudio Server. RStudio Desktop runs the program as a customary work territory application. Using the RStudio Server, RStudio running on a Linux worker can be distantly gotten to by methods for a web program. RStudio empowers customers to manage different working vaults using adventures.

It moreover has expansive group headway instruments experimental results



II. LITERATURE REVIEW

To estimate stock returns, researchers and specialists depend upon principal examination and specialized investigation. The creator [Suresh A.S] [1] portrays principal examination as the e xamination of fundamental powers that influence the prosperity of the economy. Essential investigation consolidates monetary, industry and friends examination to determine a stocks reasonable worth known as characteristic worth. As per basic examination on the off chance that the reasonable worth isn't equivalent to the present stock value, at that point the stock is either underestimated or exaggerated.

Basic examination considers macroeconomic components and individual explicit variables. Essential analys is accepted to be successful foreseeing long haul patterns. A similar paper portrays specialized examination as an enhancement for to crucial investigation yet progressively centered around predicating the cost of a security. Specialized investigation considers the adjustment sought after and supply of protections as a component of time. Subsequently it is favored over major investigation for present moment and medium term anticipating.

Specialized investigation is characterized as the craftsmanship and study of anticipating future costs dependent on the examination of past value developments by the creator [C.Boobalan][2]. Notwithstanding past stock costs, specialized investigation likewise considers organization essentials, more extensive monetary elements, advertise brain research and costs them into the stock. There are diverse specialized factors that effect and set stock costs. These indicators can be utilized for determining a macroeconomic time arrangement variable as done by creators [James H Stock][Mark W Watson] [3]. Files developed by head part examination head segment investigation are utilized to consolidate these anticipating factors. The creators built up a rough unique factor model for estimation of indexes and development of gauges. The model established a lot of 215 indicators that were mimicked in genuine time for the period 1970-1988. The strategy can be utilized to build 6, 12 or two year estimates. It was seen that during the example time frame, the given arrangement of elements gave a gauge that outflanked and other estimating strategies like Linear Discriminant analysis, Quadratic Discriminant analysis and Neural Networks. The creator additionally portrays the various difficulties that dissecting Twitter opinion presents. The absolute first challenge is looking for the correct tweets without getting excessively self-assertive. Scanning for catchphrases and deciphering slang language is another inalienable test. For powerful outcomes, the framework should be prepared on significantly more information over a bigger timeframe. A generally new strategy, Approximation and Prediction of Stock Time arrangement information (APST) has been proposed by creators [Vishwanath R.H.], et al. [6]. Experimental results show that the normal Mean Error Relative and normal Mean Absolute Error for APST are 5.90% and 0.37%. Indexes constructed by principal component analysis principal component analysis are used to combine these predicting factors. The authors developed an approximate dynamic factor model for estimation of indexes and construction of forecasts. The model constituted a set of 215 predictors that were simulated in real time for the period 1970-1988. The method can be used to construct 6, 12 or 24 month forecasts. It was observed that during the sample period, the given set of factors provided a forecast that outperformed univariate and small vector auto-regressions. The forecasts outperformed leading indicator models as well. Authors [Wei Huang][YoshiteruNakamori][Shou-Yang Wang][4] forecast stock market movement direction with support vector machine, a machine learning technique that analyzes data for classification and regression analysis. The authors investigate the predictability of the SVM technique by forecasting the weekly movement of NIKKEI 225 index. According to the paper, the key property of SVM is that training data in SVM model is equivalent to solving quadratic programming problem with linear limits. Therefore SVM always provides a solution that is unique and globally optimal. The authors also compare the SVM model with other forecasting methods like Linear Discriminant analysis, Quadratic Discriminant analysis and Neural Networks. The experimental results of the paper show that SVM has the highest forecasting accuracy as it minimizes structural risk. The integration of SVM with other methods improves forecasting performance. Author [LinhaoZhang][5] describes the effect of public sentiment on stock prices by analyzing Twitter messages.

This would allow investors to make profitable investments. Authors [JingtaoYao] [Hean-Lee POH] haves used artificial neural networks (ANN) to forecast indices of the Kuala Lumpur Stock Exchange (KLSE). Artificial neural network shave been effectively used to decode non-linear time series data. Artificial neural networks can recognize patterns and infer solutions from unknown data, thus making them extremely popular.

III. SYSTEM ANALYSIS

A. Problem Explanation

The money related market or securities exchange is unpredictable and developmental. It works as a non-direct powerful framework. As indicated by scholastic examinations developments in market costs are not arbitrary and rely on various components that associate it with present and authentic stock information. It isn't feasible for each speculator to understand the different variables that reason the costs to change. Subsequently every speculator wants a framework to foresee the future stock costs to assist them with taking proper choices.

B. Existing Frameworks

NVarious subjective and subjective investigation techniques have been created to gauge stock patterns. There are different measurable models for determining stocks and choose the opportune time to sell or hold a stock. Contingent on the organization of the information, a specific estimating model can be utilized by the speculator to foresee patterns.

C. Proposed Study

The paper proposes a modela model for anticipating time arrangement securities exchange information. The model dependent on specialized investigation utilizing ARIMA expects to mechanize the procedure of progress of stock value records. With the assistance of Information Mining systems an expectation model is created. R programing language in RStudio IDE is utilized for imagining the

II. IMPLEMENTATION

Data-mining is utilized to find designs in enormous informational collections and has wide application s in the field of measurements. Information mining procedures are concocted to address estimating issues by furnishing a solid model with information mining highlights. We utilize the auto-backward coordinated moving normal (ARIMA) model to foresee the market patterns. The total engineering of the framework is demonstrated as follows.

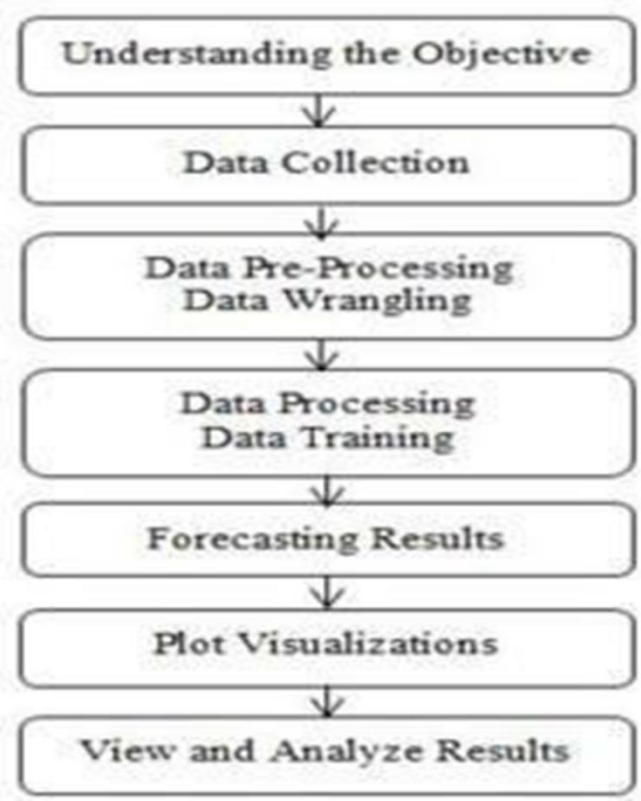


Figure.1. Implementation

Framework engineering contains the data with respect to the constituent components of a framework. It additionally portrays the connection between these components. It is a model that gives data about the conduct of a framework by breaking it into subordinate frameworks that play out similar capacities. The ARIMA framework incorporates seven significant strides to actualize the framework and each progression is explained underneath.

A. Understanding the Goal

The goal depicts the basic necessities of the framework. It helps in better comprehension of the issue explanation just as the expected results. The target this paper is to build up a framework that can be utilized by financial specialists to discover the course of the market patterns and settle on right speculation choices. The experimental results are given in a graphical organization to better translation

B. Data Collection

Understanding the target likewise helps in examining the privilege datasets. Information accumulation includes gathering data pertinent to the necessary factors and estimating them to assess results. The paper utilizes R content to gather information from Google utilizing the capacity get Symbols() accessible in the QuantMod bundle.

QuantMod

Quantmod alludes to Quantitative Monetary Demonstrating and Exchanging System for R. It is quantitative instrument that helps merchants in creating and testing exchange based factual models. The quantmod bundle makes displaying simpler and quicker by excluding rehashed work process. The bundle comprises of thorough instruments for information the executives and perception. To extract and load the information from various sources we utilize a strategy called get Symbols (). As a hotspot for acquiring the financial exchange information, the vast majority of the stock speculators use Google fund or Yippee finance. In our venture the OHLC information isn't legitimately downloaded from the Google money (finance.google.com), or Hurray finance (finance.yahoo.com) rather a call to getSymbols() is utilized to bring information. We didn't indicate the source here so the information is downloaded from default reference i.e.:- www.finance.yahoo.com.

C. Data Pre-processing:

Information gathering is approximately controlled and more than frequently trash esteems get added to the dataset. A high grouping of repetitive data (commotion) makes the information unessential and pointless for further handling. Henceforth pre-handling of information is important to set up the last dataset from given crude data. The technique

portrayed in this paper changes over the information into a separated vector list. The capacity c{base} is utilized to address the joined vector list.

- **Data Frames**

A data.frame() object in R has same dimensional properties as a framework. Be that as it may, in contrast to frameworks, information edges may contain both all out and numeric information. It tends to be said that information edge is a rundown of factors with parts as segments of a table. A rundown of factors with same number of columns and particular line names of a class is characterized as an information outline.

- **Data Processing:**

The first step in quite a while preparing is to prepare the information. The ARIMA(p, d ,q) model is utilized to process information. Financial specialists and experts two techniques to anticipate stocks to be specific auto relapse and moving normal. R gives auto.arima () strategy to estimate the time arrangement information as per ARIMA (p, d, q). The ARIMA model is an apparatus for specialized examination. It centers around rehashed parameter estimation and anticipating to locate the correct approximation model.

- **Auto Regression(AR)**

Auto regression strategy gauges the future qualities dependent on the past qualities. The capacity of an autoregressive model is indicated by AR(p), where p speaks to the request for the model. AR(0), the easiest procedure, includes no reliance between terms, going before or current. For a f irst request autoregressive model AR(1), the first term and a level of mistake add to the yield. AR(2) model considers 2 going before qualities and clamor to foresee the yield.

- **Moving Average (MA)**

A moving normal is a system to show datasets that differ as indicated by single factor. It finds the future t severs dependent on the past qualities that don't pursue a conclusive example. The two normally utilized moving normal strategies are exponential moving normal (EMA) and the basic moving normal (SMA).

- **Order of ARIMA**

The order of an ARIMA model is generally represented as ARIMA(p,d,q), where- p = order of the autoregressive part. d = degree of first differencing involved. q = order of the moving average part.

Here if d=0, then the model becomes ARMA which is linear stationary model. The same stationary and in-variability conditions that are used for autoregressive and moving average models apply to this ARIMA (p,d,q) model. Selecting the appropriate values for p, d and q can be challenging. The auto.arima ( ) function in R will do it automatically.

- **Model Estimation for ARIMA**

Model estimation for ARIMA can be achieved based on the pre-processed historical data.

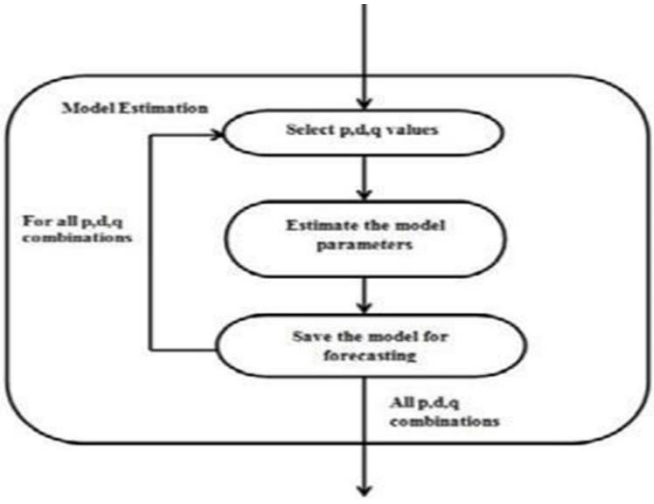


Figure.2. pre-processed historical data.

In ARIMA model, the distinguishing proof is to be cultivated utilizing auto co-connection capacity and incomplete auto co- connection work so as to

recognize p, d and q measures. For any reasonable time succession for the most part p, d and q esteems change somewhere in the range of 0 and 2, however model estimation is executed for every single likely blend of p, d and q esteems. The pictorial portrayal of these means is appeared in Fig 4.2

- **ARIMA() Function in R**

Foreseeing the correct qualities for p, dand q for ARIMA model can be extreme. The issue turns out to be increasingly unmistakable when the given dataset is bigger and contains information for a more drawn out timeframe. The auto. arima() work gave in the conjecture bundle to R mechanizes the way toward finding the correct blend of p, d and q. The estimation of d likewise affects the expectation interims i.e., the more mind boggling the estimation of d, the more quickly determining interims flood in size. For d=0, the long haul expectation normal abnormality will go to the regular aberrance of the noteworthy information. In some cases autocorrelation work (ACF) and fractional autocorrelation work (PACF) are utilized to decide the quantity ofororder of AR or MA terms required.

D. Forecasting Results

Forecasting allows us to predict future values based upon the knowledge of current and historical stock data. The model specified here uses the forecast package for R for predicting future stock values. The forecast package contains tools for analyzing univariate time series data using state space models and ARIMA modelling. The Arima () and auto. arima () functions used to model future stock prices are a part of the forecast package.

E. Plot Visualisation

Plot representation includes speaking to the numerical information in graphical configuration. In the given approach, line diagrams and histograms are utilized to speak to the stock information. This is finished utilizing the plot () capacity gave in R. The include BBands () capacity includes two extra lines that make information understanding simpler. The x-pivot speaks to the speaks to time span as far as year/months and days while the y hub shows stock value esteems.

III. MODEL SIMULATION

The step by step execution and code is provided below. We will start with the same basics of running basic checks on the data and then take a deeper dive in terms of modelling technique to use.

IV. CONCLUSION

In this paper an undertaking was made to check the monetary trade expenses of the MICROSOFT stock by working up a desire model subject to particular assessment of evident time course of action data and data mining methods. This paper succesfully foreseen the stock worth records for flashing period using an ARIMA model. The capacity of the ARIMA model in finding future stock worth records which will enable stock operators/theorists to make beneficial endeavor is tremendous. The simply burden of this model when contrasted with its adversaries is the penchant to handle the mean of the chronicled data as gauge concerning long stretch expectation. Accordingly it isn't judicious to use this model for long stretch deciding of stock worth records.

V. FUTURE SCOPE

The possibility of integrating this model with fundamental analysis can lead to better decision making when it comes to making decisions like buy/hold/sell a stock. Through a pertinent sentiment analysis performed by collecting social media data and combining it with the ARIMA forecast better profitable investment decisions could be made.

VI. REFERENCES

[1]. Fayyed, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, AI Magazine 96, 37–54 (Fall 1996)

[2]. Fiol-Roig, G.: UIB-IK: A Computer System for Decision Trees Induction. In: Raś, Z.W., Skowron, A. (eds.) ISMIS 1999. LNCS, vol. 1609, pp. 601–611. Springer, Heidelberg (1999)

[3]. Weinstein, S.: Stan's Weinstein's Secrets For Profiting in Bull and Bear Markets. McGraw-Hill, New York (1988)

[4]. The R venture for Statistical Computing, http://www.r- project.org/

[5]. http://www.nytimes.com/2009/01/07/innovation/ business- figuring/07program.html

[6]. Miró-Julià, M.: Knowledge Discovery in Databases Using Multivalued Array Algebra. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) Computer Aided Systems Theory - EUROCAST 2009. LNCS, vol. 5717, pp. 17–24.

[7]. Fiol-Roig, G.: Learning from Incompletely Specified Object Attribute Tables with Continuous Attributes. Boondocks in Artificial Intelligence and Applications, vol. 113, pp. 145–152 (2004)

[8]. Quinlan, J.R.: Induction of choice trees. AI 1, 81–106 (1986)

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Neuromorphic Computing

Dr. Avinash S. Kapse[1], Ms. Sandhya A. Kale[2], Ms. Asha Gaurave[2]
[1]HOD, Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India
[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

This paper gives an outline of the difficulties looked by equipment executed Spiking Neural Networks, from gadget to circuit plan, unwavering quality and test. We present a far-reaching depiction of the best-in class neuromorphic models enlivened by cerebrum calculation, with extraordinary accentuation on Spiking Neural Networks (SNNs), along with arising advancements that have empowered such frameworks, specifically Stage Change and Metal Oxide Resistive Memories. At long last, we examine the principle challenges looked by equipment usage of SNNs, their unwavering quality and post-creation test issues

Keywords — Spiking Neural Networks (SNNs), Phase-Change Memories (PCMs)

I. INTRODUCTION

Equipment execution of neural organizations is a hot research subject and is currently considered as key for a few huge equipment situated organizations, for example, Nvidia, IBM, Intel, just as programming focused organizations, for example, Amazon, Facebook, Microsoft. The new interest around profound neural networks for design acknowledgment and characterization has put a new focus on neuromorphic registering that brings cerebrum demonstrating nearer to information examination. Top undertakings in neuromorphic designing have prompted ground-breaking mind motivated chips ready to mimic various spiking neurons working in non-von Neumann PC designs. These innovations need to fit inserted frameworks or Internet-of-Things (IoT) necessities consequently, their energy utilization is basic also, should be limited. Heterogeneous mix among CMOS and rising advancements is viewed as an freedom to achieve such objective. Without a doubt, arising innovations have the capability of giving numerous advantages, for example, energy effectiveness, high mix thickness, CMOS compatibility, reconfigurability, non-unpredictability, and open the way towards novel computational designs and approaches, for the customary Von-Neumann structures and past. Among the arising advances, memory advances such as Resistive Memories (ReRAMs), Phase-Change Memories (PCMs), or spintronic based recollections (STT-MRAMs) are setting off exceptional interdisciplinary movement, having driven the research local area towards returning to the current registering also, capacity standards, giving equipment answers for neuromorphic registering. Considering the huge number of neurons and neurotransmitters needed to perform productive learning what's more, characterization, plan groups face a few deterrents: productive capacity of the synaptic loads, admittance to boundaries progressively, dependable and testable plan of cross breed, analogdigital- Non-Volatile

heterogeneous structures. Hence, it is critical to mutually think about gadget material science, circuit vigorous plan, testability, and engineering limitations by and large. The reason for this paper is to give a complete outline of the bio propelled equipment executed neuromorphic structures, with uncommon accentuation on their plan, unwavering quality and test. The paper is coordinated as follows. The subsequent segment presents a thorough depiction of cutting-edge neuromorphic models propelled by cerebrum calculation. The third area depicts the most recent accomplishments in Phase Change Memory and Metal Oxide Resistive Memory arising advancements utilized as counterfeit neurotransmitters for Spiking Neural Networks (SNNs). The fourth segment sums up the fundamental imperatives of equipment executed SNNs, potential shortcomings and unwavering quality issues, along with primary difficulties looked by post-creation test

II. NEUROMORPHIC ARCHITECTURES: A BRAIDGE ARCHITECTURE: A BRIDGE BETWEEN MACHINE LEARNING AND BRAIN- INSPIRED COMPUTING

Counterfeit neural organizations, enlivened by the gigantic capacities of the organic cerebrum, have been continually returned to, routinely opening new fields of utilization (PC vision, AI, advanced mechanics, counterfeit knowledge, and so forth) A few electronic substrates are presently examined that offer intriguing attributes to accomplish, on the one hand, the guarantees of energy proficiency of the natural model and, then again, the innovative development and the programming limits expected by the applications. The principal challenge for neuromorphic equipment configuration is identified with their force productivity during the deduction period of the neural organization in the application areas identified with implanted frameworks. Learning is then expected to be finished disconnected and isn't considered in the difficult definition. In this setting, numerous equipment quickening agents have been proposed as of late, (for example, Many center, FPGA and SIMD) and they are presently completely incorporated in Machine Learning (ML) systems what's more, accessible as business arrangements. Nonetheless, equipment Spiking Neural Networks (SNN) have appeared to bring an improvement in proficiency contrasted with old style activitybased Counterfeit Neural Networks (ANN) while guaranteeing practically identical order exhibitions [1]. SNN- based equipment stays an exceptional subject of interdisciplinary research since it investigates the conduct and the incredible capacities of the cerebrum. Albeit some encouraging arrangements as of now exist [2], they just translate the old style ML approaches into another coding plan. The worldview change from ANN to SNN would truly be arrived at when the properties of bioinspired neurons will be uncovered. Undoubtedly, natural properties of bio-enlivened neural organizations show unaided and dispersed learning contrasted with ML approaches. In any case, exploiting these properties requires a superior comprehension of the worldwide conduct of the cerebrum through interdisciplinary exploration from neurosciences to hardware. This change in perspective starts with the scrutinizing of the back-engendering calculation which shows each day its limits regarding energy utilization and versatility. This concentrated strategy playing out the calculation of order mistake appears probably not going to be actualized by the mind, which plainly shows the interest and advantage of neighborhood, conveyed and unaided figuring. The related bio-propelled learning rule is currently known as STDP (Spike Based Dependent Plasticity) and is applied on every neural connection freely of the worldwide condition of the organization. Consequently, the neural connection should be gushed of calculation capacities. Quite possibly the

most encouraging arrangements is to utilize resistive materials, as the ones depicted in area III. Moreover, the difference in learning rule likewise raises the question of the feedforward geography itself as we discover it today in current profound organizations. As a matter of fact, STDP is a sort of serious learning measure, where every neuron learns and addresses a prototypical info. The opposition between neurons is guaranteed by horizontal restraint, indeed more naturally plausible. It is executed with a WTA (Winner- Takes-All) all- to-all availability to guarantee the combination of the learning and the great portrayal of the input information [3]. Such a geography unmistakably changes the equipment usage yet additionally acquires excess the portrayal of classes of information which consequently is likewise more dependable face to equipment issues or fluctuation saw in the latest plan measure (talked about in segment IV). At long last, such solo learning changes the current perspective about the utilization of learning techniques by tending to the issue of on-line learning, or even long haul learning, normally present in natural frameworks. It clears the approach to new application areas where the framework can adjust persistently to its current circumstance. Tending to together all the specificities of bio-enlivened learning could give a response to the issue of extreme utilization of current disconnected learning procedures. The principle challenge of neuromorphic equipment is identified with the control of the learning interaction itself. It stays still hard to adjust it to various applications, moreover while considering unaided learning. This additionally brings up many energizing issues also, challenges in the field of microelectronic plan, for example, joint plan with heterogeneous innovations, simple/advanced interfaces, programmability of neuromorphic circuits and materials, dynamic administration of their utilization, interoperability with advanced applications, lastly the confirmation and unwavering quality of these frameworks

III. METAL OXIDE RESISTIVE MEMORY AS ARTIFICIAL SYNAPSES IN SPIKING NEURAL NETWORKS

Spike based computational instrument and design co-restriction of preparing and memory are two significant highlights to plan and create equipment SNNs. This requires the reconciliation of equipment neuron circuits must with particular circuits demonstrating neurotransmitters. Neurotransmitters need to show pliancy, that is regulation in their adequacy, and to uphold web based learning calculations, that show in changes in their conductivity. Stage Change Memory (PCM) and Metal Oxide Resistive Memory (OxRAM) can be utilized as synaptic components on account of their tunable conductivity, similarity with cutting edge CMOS creation measure, low force utilization, non unpredictability and adaptability. Two principle ways to deal with imitate synaptic conductance regulation have been effectively illustrated: a) the simple methodology, where different opposition states to imitate long haul potentiation furthermore, discouragement (total diminishing and increment of opposition, LTP and LTD) are utilized; b) the parallel methodology, where just two unmistakable obstruction states (Low Resistance State, LRS, and High Resistance State, HRS) per gadget related with a probabilistic programming plan are received. In the last case, the conductance adjustment is accomplished by planning a solitary neural connection as a sythesis of n numerous twofold cells working in equal [4]. PCM gadgets can be utilized as a simple memory. In any case, they show a solid imbalance between the SET and RESET measure: the SET cycle is amazingly slow, fundamentally the same as synaptic potentiation, while the RESET cycle is unexpected. In [5] a 2-PCM neural connection that reproduces fake evenness among SET and REST by utilizing two gadgets for each neural connection has been proposed.

This technique functions admirably, yet it requires long and eager for energy revive activities, where all PCM gadgets are being reconstructed. A thin warmer base cathode based PCM related with an introduction electrical heartbeat followed by an arrangement of indistinguishable quick programming beats is introduced in [6] to execute a bidirectional neural connection. Continuous long haul potentiation and sorrow are accomplished by applying a long train of indistinguishable short SET and REST beats (< 50 ns). The outcomes are represented in Figure 1. Under these conditions, the made undefined district doesn't cover the whole territory of the base anode (for example the radiator), accomplishing a steady opposition tweak. This technique has been exhibited to accomplish unaided getting the hang of utilizing STDP, on the character acknowledgment application. A normal characterization pace of about 76% has been exhibited by methods for framework level reproductions. In concurrence with test results, 200 SET furthermore, 30 RESET levels have been mimicked. To improve the execution, the quantity of misery levels must be expanded. A characterization pace of 82% is accomplished for 100 misery levels (i.e., 100 RESET).

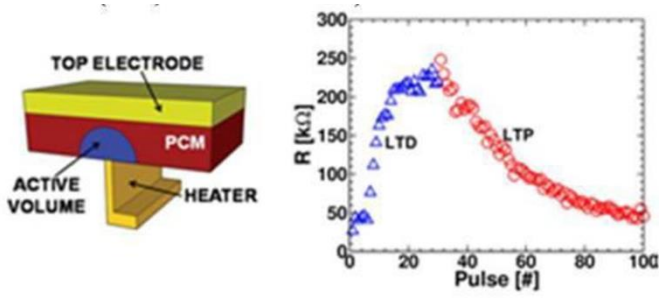


Fig. 1. Phase Change Memory (a) Schematic drawing of the PCM wall storage element. (b) Depression (LTD) and potentiation (LTP) characteristics. The memory is initialized by applying a 20 ns pulse of 1.25 V to set the cell at 30 kΩ. A train of identical 20 ns amorphizing pulses of 1.6 V (LTD) and 20 ns crystallization pulses of 1.25 V (LTP) are applied. OxRAMs can be customized devouring little power and can be incorporated essentially with cutting edge CMOS innovations. Be that as it may, these gadgets experience the ill effects of high cycle-to-cycle and gadget to-gadget conductance inconstancy, introducing extensive test for standard memory applications. Figure 2 shows the total dispersion of HRS and LRS for a TiN/HfO2/Ti/TiN stack. The estimations have been performed on a 4 kbit exhibit. In OxRAMs, the two bearings of writing computer programs are non- aggregate. Consequently, an elective strategy to execute stochastic STDP has been proposed [7]. The proposed learning plan has been exhibited on a Fully Connected Neural Network for car following. It has been shown that the organization can draw profit by conductance fluctuation since it expands the scope of synaptic loads accessible during learning [8].

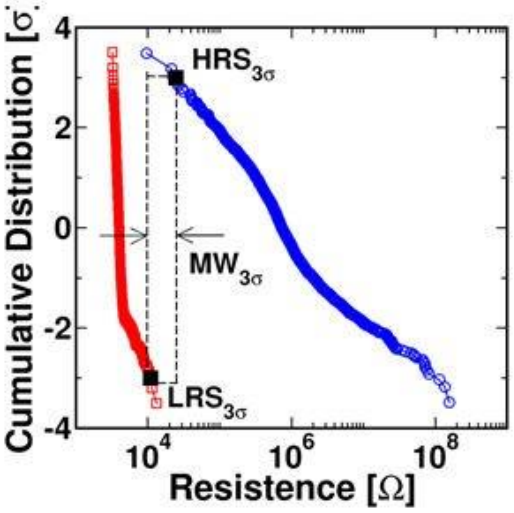


Fig. 2. Cumulative distribution of LRS and HRS measured on 4 kbit array (Vset = 2 V, Vreset = 2.5 V, Icc=200μA).

IV. ON THE DESIGN, ROBUSTNESS AND TEST OF SPIKING NEURAL NERWORKS

To get the greatest proficiency out of equipment executed SNNs, useful modules (neuron-neurotransmitter) must be planned so that their info/yield qualities give the learning and preparing ability needed by application. Furthermore, the

organization availability needs to take into account high combination with solid and dependable reconfiguration and transformation attributes. Critical advantages can be acquired by receiving RRAMs for neuromorphic calculation as clarified in the past area. RRAM with bidirectional and consistent conductance tuning ability is considered as a characteristic electricallycontrolled synaptic gadget. There is a lot of work committed to the advancement of new synaptic-agreeable gadgets, gadget displaying and algorithmic approval of such gadgets set with regards to SNNs. Little work has been committed to the actual plan of a full SNN and the ID of potential issues and dependability issues and no post-creation test arrangements have been proposed up until this point. For a strong and proficient equipment execution of SNNs we need to together think about the qualities of the SNN itself (availability, neuronal enactment work, learning rule and synaptic update) and the attribute of the gadgets used to execute it (CMOS ON/OFF current and limit voltage, conductivity adjustment and current consistence of the RRAM, and so forth) Here we center around a completely associated SNN, that picks up utilizing the Spike Timing Dependent Plasticity (STDP) technique with horizontal restraint, with coordinate and-fire neuron and resistive neural connections (see Fig. 3). Fig. 3a) outlines the availability between two back to back layers of such an organization. Here N addresses a spiking neuron, S addresses a neurotransmitter and test covering is the hardware not-relating to the SNN however important to play out the post manufacture test. The neurotransmitter furthermore, its control circuit are represented in Fig. 3b). A few works propose utilizing the resistive-based synaptic gadgets in a crossbar exhibit (without access gadget) to guarantee least territory impression. Notwithstanding, such usage experiences enormous spillage flows in the half gotten to cells (seriously unfavourable to the learning cycle) and from wasteful shaping interaction (at the point when RRAM gadgets are utilized for neurotransmitter) and un-controlled current consistence (can cause neurotransmitter disintegration or failure). Thus, in the introduced plan, an access semiconductor is utilized for the admittance to every neural connection. This semiconductor ought to disconnect the neural connection from the remainder of the network when there is no movement on both of its associated neurons, and it ought to permit the entry signal coming from both of the neurons. This conduct is ensured by OR-ing empower signals produced by the neurons. These signs are created each time a neuron spikes. The PPre is the presynaptic spike, comes from an information neuron and caries the input data, while PPost is the postsynaptic spike, comes from a yield neuron and caries the learning control. The age of these heartbeats is finished by an incorporate and-fire neuron. One neuron gets data from all neurons in the past layer of the organization, balanced by the comparing synaptic weight. This data is collected until it arrives at a specific level, so, all in all the neuron imparts sign towards the following layer. The utilitarian design of a neuron is outlined in Fig. 3 c). The gathering capacity is performed by an integrator structure, addressed in the figure by the operational intensifier and capacitor C. The yield of this intensifier is analyzed against a limit Vth, and once the edge is accomplished, a heartbeat generator is initiated, who's principle usefulness is to produce the postsynaptic spike (the PPost beat) input to change the synaptic weight, and feedforward, as presynaptic spike (PPre beat), to send the data to the following organization layer. In expansion to this data conveying signal, the beat generator needs to give the empower signs to control (I) the switch S1 and initiate the neuron's recalcitrant period, (ii) the switch S2 and initiate the synaptic weight balance (learning), or (iii) to permit the section of the balanced presynaptic spike (W*PPre, with W being

the synaptic weight). Furthermore, the beat generator gives the empower signals controlling the OR door in the neurotransmitter control and the crippled signs restraining adjoining neurons (horizontal restraint) - not appeared in the figure. When the circuit is planned and its boundaries advanced by reproduction, it will be manufactured. Assembling cycle of coordinated circuits as a rule is not completely controlled. There are numerous impacts that could lead to imperfect circuits, including dust, spot surrenders on the silicon wafer, measure fluctuation, and gathering shortcomings. Assembling testing is the cycle ready to guarantee quality and unwavering quality of coordinated circuits. The primary methodology used to test coordinated circuits depends on: (I) the age of test input vectors ready to target potential issues; (ii) the application of those vectors to the circuit; and (iii) the correlation of the reactions gave by the circuit the normal ones recalculated by a test system. While numerous effective arrangements exist for testing conventional plans (simple, computerized or blended sign), to the most amazing aspect our insight, there is no work so far managing with the post-creation testing of an equipment executed SNN. As portrayed beforehand, circuits actualizing SNNs have some significant contrasts contrasted with old style circuits. To be sure, they resort to the blend of gadgets with both deterministic and stochastic practices and they incorporate both computerized and simple components. A test system appropriate for SNNs ought to have the option to test the right activity of the two neurons and neural connections, it ought to be prudent (in region, power and execution overhead), it ought not rely upon the preparation information nor on the setting the organization would be utilized. This last condition is significant for broadly useful SNNs, which could be utilized in various situations. In fact, it is workable for SNNs with web based figuring out how to be utilized for design arrangement inside various settings, as long as the issue space is equivalent or more modest than the info neuron space. To distinguish and execute a productive test system, it is important to comprehend the conceivable broken conduct of the circuit under test. This is generally achieved by flaw demonstrating. The flaws that can be found in resistive memory exhibits can be ordered into two classifications: delicate shortcomings and hard blames [9-10]. Delicate flaws are brought about by various cycle- to-cycle or gadget to gadget varieties that show up during the creation, yet additionally in-field during read/compose activities. Hard blames are incited by creation steps or they can be brought about by the shaping interaction or on the other hand by ceaseless pressure; they are more hard to be forestalled. One run of the mill kind of hard deficiency happens when the obstruction of a resistive memory cell will at this point don't change; this class incorporates stuck-at-0 (SA0) and stuck- at-1 (SA1) flaws brought about by creation strategies and restricted perseverance. For this situation, the flawed gadget is stuck at high opposition or low obstruction state, and these circumstances are happening with a very high likelihood. It is accounted for that 63% of a capacity exhibit in light of memristor are without issue in a 4Mb resistive RAM, with about 10% of the cell being of Stuck-At type. In the creators showed that 10% of broken RRAM cells will lead to significant corruption of the precision and generally execution of a convolutional neural organization. These investigations were performed for resistive gadgets utilized in memory clusters, where the resistive states are viewed as parallel. Nonetheless, restricted research has been devoted to examine and display the issues of a resistive neural connection, where the resistive gadget is either utilized as a simple gadget (numerous obstruction states are utilized to copy the synaptic loads) or as a parallel gadget yet with a probabilistic programming plan. Along these lines, deficiency models what's

more, test procedures must be concocted by the explicitness of the synaptic cluster. For the simple neurotransmitter usage, a traditional simple test could be utilized, while for the paired execution with probabilistic programing a completely new test technique ought to be concocted. The fundamental challenge here is to distinguish a procedure to test for the arbitrariness of the resistive gadget programing. This test has to be minimal effort and no affect the organization conduct at runtime. A chance is to begin utilize comparative procedure as the ones utilized for the testing nature of genuine irregular number generators (TRNGs). Nonetheless, since for this situation the entropy isn't of significant concern, minimal effort test, for example, NIST , could be utilized. Concerning neuron, it is much of the time an simple gadget, and traditional simple test can be utilized. The primary challenge here is to figure out how to apply the test vectors and peruse the circuit reaction. Despite the fact that trustworthiness investigation and plan for dependability are presence of mind when managing customary (Von Neumann) figuring structures, they are not really basic when managing spiking neuromorphic structures. It was viewed as that neural organizations are characteristically tough as they infer algorithmic strength, regularization, over parameterized information. Be that as it may, when the organization is actualized in equipment, the adaptation to non-critical failure property debases because of inserted frameworks limitations, for example, solid restricted size of the organization (i.e., number of shrouded layers and number of neurons in each layer) and more modest neuron network (i.e., number of neural connections), size decrease of loads. Because of the way that we target least size organizationsfor most extreme (conceivable) intricacy and low force, equipment flaws become important. Papers show primer aftereffects of innovation steadfastness dangers sway on neuron and synaptic usefulness, just as the measurement of their consequences for the activity proficiency of the useful decreased size neurotransmitters and furthermore spiking neural networks. Assessments have been performed on ordinary 2- layers of completely associated spiking neurons with horizontal restraint. STDP is utilized as solo learning for the acknowledgment of static pictures and the acknowledgment mistake can drop by over 35% with a diminished arrangement of preparing information. In bigger neural connection stockpiling abilities, the drop of mistake acknowledgment is as yet adjoining 20%, and can be brought to 10% just if the preparing set is expanded by at any rate multiple times, which thusly will produce comparative expansion in the force envelope

V. CONCLUSION

Another plan rule of neuromorphic figuring framework based on the current discoveries in mind science was proposed and the 'Tangji' chip was planned and created dependent on it. A multi-chip design-based PCB board was executed. The comparing programming framework and reproduction stage were created for additional applications on the chip.

VI. REFERENCES

[1]. L. Khacef, N. Abderrahmane, B. Miramond, “Confronting machinelearning with neuroscience for neuromorphic architectures design,” International Joint Conference on Neural Networks (IJCNN), 2018

[2]. L.A. Camuñas-Mesa, et al.,“A Configurable Event-Driven Convolutional Node with Rate Saturation Mechanism for Modular ConvNet Systems Implementation,” Frontiers in Neuroscience, 2018

[3]. R. Kreiser, T. Moraitis, Y. Sandamirskaya and G. Indiveri, “On-chip unsupervised learning in

winner-take-all networks of spiking neurons," IEEE Biomedical Circuits and Systems Conference (BioCAS), 2017

[4]. D. Garbin, et al., "Hfo2-based oxram devices as synapses for convolutional neural networks," IEEE Transactions on Electron Devices, vol. 62, no. 8, pp. 2494–2501, 2015.

[5]. M. Suri, et al. "Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction," 2011 International Electron Devices Meeting, 4.4.1- 4.4.4.

[6]. S. La Barbera, et al., "Narrow Heater Bottom Electrode-based Phase Change Memory as a Bidirectional Artificial Synapse," to appear on Advanced Electronic Materials.

[7]. M. Suri, et al. "CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (cochlea) and visual (retina) cognitive processing applications," 2012 IEDM.

[8]. D. R. B. Ly; et al., "Role of synaptic variability in spike-based neuromorphic circuits with unsupervised learning", ISCAS 2018.

[9]. (Degrae2015) R.Degraeveetal.,"Causes and consequences of thestochastic aspect of filamentary RRAM," Microelectronic Engineering, vol. 147, pp. 171–175, 2015

[10]. E. I. Vatajelu, P. Prinetto, M. Taouil, S. Hamdioui, "Challenges and Solutions in Emerging Memory Testing", IEEE TETC, 2017

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Analysis of Effective Approaches for Legal Texts Summarization Using Deep Learning

Ms. Shruti J. Sapra (Thakur)[1], Dr. Avinash S. Kapse[2]

[1]Research Scholar, Department of Computer Science and Engineering, Amravati University, Assistant Professor, Department of CSE, G H Raisoni College of Engineering, Nagpur, Maharashtra, India

[2]HOD, Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

ABSTRACT

Now a day there is a great importance is associated with the text summarization as there is a large amount of data is available on the different social media platforms, websites and blogs. Also there are large number of tools used for the summarization of the text available in the different forms this text which is available from the different sources are present in the various format some of the data is in the structured and some are in the unstructured format. The user or the application which is using this data has to be careful about the use of the data as the data is available has to be authentic and must be delivered to the user in the proper format. The large amount of time is associated with understanding the meaning of the data. Therefore user has to invest large amount of time in reading the data in other words there is a great deal in getting the meaning of the data as the data is in large amount and hence getting the meaning from such a huge amount of data is not a simple job. The best option is to summarize means to shorten the data in so that user will require less amount of time in understanding the meaning of the data and will be useful for the user or the application in which it is used.

This paper essentially discusses the various approaches from the domain of the deep learning and machine learning in shortening the data as well as summarizing the data. Different algorithms are also discussed in order to understand the various parameters associated with the text summarization. Also the best algorithm must be chosen for the effective text summarization process for the better results and better efficiency. Also in India there is a great importance is associated with the legal text but many a times the common man or the user is unable to recognize the meaning of the text by means of the text is very large and descriptive hence it is very essential to decode such a text in simple and short text. Thus the specialized algorithms are used to interpret the meaning or directly the text in to the format the use is able to understand. In other words the techniques have been discussed shortens or summarizes the legal text and makes easier or readable for the application or the user.

Keywords: Deep learning algorithms, legal text summarization, Abstractive text summarization, strategies for summarization.

I. INTRODUCTION

Text summarization is the restating of the actual information of the text and make that text as short as possible or in other words expresses the large information in a very few words or sentences as it is possible. [1] This process will significantly reduce the size of the text and but the information should

convey the same meaning as previously conveyed. This increases the readability of the text for the user of the application. [2] And this task can be done using the various tool such as artificial intelligence machine learning also with the support of dissimilar deep learning algorithms available in the previous research [3]

Text summarization also referred as the unique the method of shortening very long words, sentences or paragraphs of text or data. The clear intention behind this is to make coherent and short summary having the important notes which is short text or data which is used in number of applications for the purpose of proper understanding in small amount of time. [4]

Many times the text summarization is Automatic in nature and is a major issue in medical field and hence different techniques has to be applied for such a type of data which is very crucial. [4]

Summarization can also be defined as reduction in the major content of the information whereas the original meaning of the information. Has to be there in the short text .which will be conveyed to the intended user of the information.

Most of the author has worked on the Abstractive summarization method which is the unique method of summarization of the text and essentially very useful considering the different domain of the application and has number of advantages over the other techniques discussed [5], On the other hand is a method where the short summary is produced by creating new paragraphs by either replacing or reconstructing the paragraphs, instead of making use of different use of words having the same meaning and the important paragraphs.[6]

Legal text can be various kinds of information written for different purposes related to the law, including: book about law. Legal treatise, Law book, a major publications related to all the law or related to field of law. In other words, texts created in the course of legal research. [7]

II. BACKGROUND

Different researchers from the different areas studied different aspects of the text summarization specially the text summarization in concerning with the legal text. They found the various parameters which affect this process of summarization and also used different methods for the text summarization.

Summarization of Text is the complete process of filtering the significant information from different sources like internet and others to create a unique version for a particular applications or user and work. [7]

There are mainly two types of summarization of the text in this regard which are discussed as below

a. Abstractive Summarization

Abstractive summarizers are one of the efficient methods of summarization in which sentences are not selected from the main given information or passage to create the summary. Instead, they may create a paraphrasing of the related main information of the given information, using a different set of vocabulary from the main information.

b. Extractive summarization

Extractive summarization means identifying significant sections of the information and producing then producing a unique subset of the paragraphs from the main information; while abstractive summarization may produce matter in a completely new way after completion of this process of summarization of the information using various natural language processing tools

From the important findings from the previous research it is found that the abstractive summarizer performs more better over all the different methods of the summarization in concerning with the different parameters of the

summarization, the outcomes recommend that the margin by which abstraction outperforms extraction is greater when controversialist is high, providing a context in which the need for generation- based methods is especially great.

There is another approach which is used in the text summarization i.e. Latent Semantic Analysis which is an unsupervised learning algorithm and thus can be applied for extractive summarization of text. First, import the summarizer. [7]

The second approach used in the process of summarization is the application of different tools. These tools are used according to the application need. Some of the applications requires the basic summarization while the others requires high level of the summarization of the text. The advanced level of summarization requires the high level of tools which are very efficient in summarization and also requires very less amount of time for the complete process of summarization on the other hand the basic tools are not that much efficient as compared to the advanced tools but are more cost efficient as compared to the advanced tools. Thus the user will select according to their preferences which tool is too used for the summarization. In case of basic tools there may be an error in summarizing the text or information and May sometimes leads to wrong information in the summarized text. [8]

Therefore it is always a better choice to use the advanced tools for thee summarization. These tools involve the application of advanced algorithms hence they are more effective as compared to the basic algorithms. These advanced algorithms are the nothing but deep learning algorithms also advanced machine learning algorithms.

The paper is arranged in following manner:

Section I Introduction. Section II describes Background. After that Section III describes previous work done by various researchers in this domain. Section IV describes various existing methodologies. Section V describes various attributes and different parameters. Section VI gives new approach. Section VII experimental tests carried out. Section VIII is approaches for Testing's, Section IX is outcome and results. Section X is conclusion. Finally Section XI is future scope of this analytical paper.

III. PREVIOUS WORK DONE

Some of the approaches use the advanced machine learning and neural network algorithms which are more efficient in summarizing the text or information. The main advantages of such algorithm are they require less amount of time for the process of summarization and mainly convey the same information as the main or original text was conveying. One of the best approaches is the application of Bidirectional transformer is a best transformer applied to overcome the limitations of RNN and other neural networks as defining characteristics of Long term dependencies present in the information or the text. It is a pre-trained and well defined structure which is very efficient and is naturally bidirectional. This well defined structured approach can be tuned to easily to perform various functions of NLP tasks as specified in the different application or as per the requirement of the user which will suit the requirement of the user for the particular application.

Hence this model or structure is more likely to use in most of the summarization process and is very beneficial for the particular application point of view.

There is another approach which is used by most of the researchers in the summarizing the large amount of text and i.e. PEGASUS this is Advanced Model or structure for Abstractive and extractive Summarization of text and is also the best approach for using the different sentences or in other words replacing the same meaning sentences by another sentences or paragraphs this approach will also

significantly reduce the text to a short text and is also a better choice for the summarization of the text or information.

This approach for summarization of text is one of the most challenging and difficult task in case of legal text is involved as there is a use of very critical rather a very specific words are used in this context an many times the another word or sentence for the same word or sentence is not suitable for the replacement and in such a case the tasks can be completed with the help of specific algorithms in processing a natural language which may involve understanding of long paragraphs in legal context and find the meaning of the each sentence or words in the paragraph. Sometimes this method also use the approach known as information or data compression and producing a new language, this approach is also a typical approach and also involves many challenges which has to be solved using a specific algorithms and special techniques.[8]

IV. EXISTING METHODOLOGIES

There are different methodologies which are efficient in summarizing the text some are very advantageous and applicable in the wide range of the applications hence they are more popular in this context and also are generating effective results as compared to the other methods of summarization [8] These methods and techniques may consider different or same parameters for the summarization like accuracy performance and time required for the performance. Researchers deeply understand the phenomenon of summarization and then applied the different methods and techniques according to the application domain. [9]

Out of which some of the methods are discussed here some of the methods explained [3] gives importance to the repetitive words, sentences and paragraphs used in the summarization and also used for the specific application. Their main aim is to reduce the size of the text or information while the meaning should not change accordingly. Therefore they Include keywords and terms used by the different researcher and writers and carefully think over the text or information and also about how the sources of different ideas which are relevant to the application domain and must be suitable for the application where they are going to use. Applying only short notes, might explain the genuine writer's main concepts in different form. Then explain how those concepts for the same context.

The next approach which is used in the Google is also studied here and also helpful in understanding the minute details of the summarization of the text which is also very useful in many context. This approach explain the importance of the Not to be outdone for the particular text or information which is used in the legal context and therefore be applied for the legal text or information summarization, a Google Brain and his team in the Imperial College London has developed a complete system and such Pre-training with more and efficient extracted text in creating the Gap-sentences for Summarization of typical text like legal text and legal information like court judgments and court amendments [10] The clear sentence to sentence sequence is applied to maintain the flow of information and make the meaning clear to the user or the application

This approach also used the concepts of Pegasus in which the summarization of the legal text and then interpret in the different form is done and in which the user can understand and interpret this shot text.

This approach is also applicable for leverages Google's Transformers structure which may be combined with predefining the various objectives tailored according short text production.

V. ANALYSIS AND DISCUSSION

Analysis of the summarization process is based on the different parameters such as word count, number of

sentences reduced and number of paragraphs reduced. In addition to this there are various parameters relating to the technology are the time required for the summarization, space required for the text or information before and after the summarization of the text or information. Then there are parameters relating to the application in which these methods or techniques applied. All the approaches are studied in detail and analyzed according to the different parameters discussed above.

In addition to this there are more important parameter is the used to technology and the use of the algorithm using which the summarization of the text is done.

VI. PROPOSED METHODOLOGY

This method proposed in this section is very useful for the documents relating to the court judgments and applied in various applications relating to the legal aspect. The different head notes and sections of a various court Judgments regarding the legal document summarization use the case of text summarization but it does not relate to the paragraphs of the other judgments of the courts of the other nation.

This method comprises of two main steps, first is the Generation of better quality information by expressing the short notes present as headings of the documents of the legal notes and second is applying this information in context to extraction of the significant words or paragraphs mentioned in the document which is to be inserted. This research proposed has several benefits over the method discussed earlier. For the implementation of the algorithm. The word count table is generated for the words those are repetitive in nature in the given legal text such words are replaced by the same meaning words from the dictionary of the dataset and a new short summary is created which is very short in nature as compared to the previous original text and that also can be used in the different applications like souckal media, blogs and other websites where summarization is essential. Thus this approach maintains a special dictionary for the words in the table which are recitative and which are having the same meaning as that of the other.

After this process of word replacement sentences are Tokenize the replacement and similar or repetitive meaning sentences are replaced using a dataset. Thus the set of sentences are replaced in a sequence to form a short notes or short text. This process is continuously implemented to generate a summary which can be directly applied to the application of domain.

VII. APPROACH FOR TESTINGS

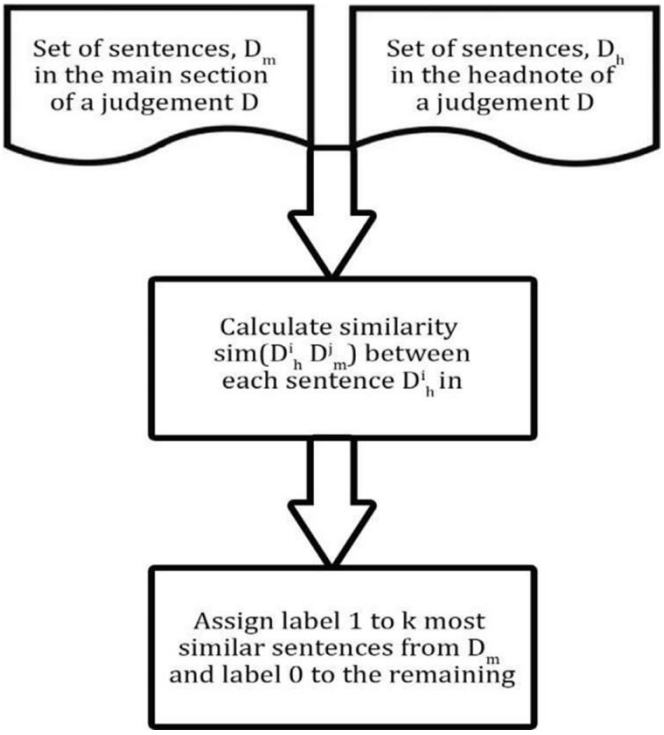


Fig 1. : Patterns and data set generation in the legal text.

Labelled dataset generation

This stage involves the production of various useful dataset for the replacement of the data from the actual information. Thus using this ready data set the

short text or information is produced and applied to the different applications. Some of the case judgments may contain short summary which is not to be replaced and must match with our dataset and this will significantly save the time for the summarization. Such an information is highlighted as not to replace. Similarity, there are some words or significant sentences which are not be replaced such words and sentences has again to be labeled for non replacement. This approach used the Text Rank which is a graph-based ranking structure. Text processing can be done in order to find the related paragraphs or related sentences in text and also to find useful words and after that replaced them continuously

Single-document is summarized using transformation of the source text into a compressed, shorter text in which the irrelevant data is deleted. Using this approach large and complex legal documents are summarized in the best short form which are easily readable and also takes less space on the disk.

VIII. IMPLEMENTATION

The implementation of the proposed method is done using a deep learning algorithms which are very efficient and follows the advanced text summarization approach in which less amount of time is needed and following procedure is applied

The proposed approach can be tested using the different tools of the summarization and results are generated which are compared against the different essential parameters and best summary is generated.

IX. OUTCOME AND RESULTS

It is very clear from the above proposed method that the different parameters studied according to different methods behave differently in every situation and hence inbound useful for this approach. This proposed method foods very useful in all scenarios and various applications where it is implemented and hence found suitable for the real time and other applications also.

X. CONCLUSION

This contribution is multidimensional approach in which this proposed approach a new technique of generating a unique summary which is the need for expert summarization. This approach finds the best in such a complex domain in which other techniques of summarization may not work efficiently. This approach finds the most suitable technique in the field of summarization of the text or information from the legal documents.

XI. FUTURE SCOPES

It is expected that the continuous research and development will eventually result in a variety of utility and tools. These strategies will also improve the effectiveness and efficiency for implementing various methods and technologies of text summarization of legal information. So, overall enhanced summarization approach will improve the proposed method significantly.

XII.REFERENCES

[1]. A full Python Implementation of the ROUGE Metric (not a wrapper). Retrieved from https://github.com/pltrdy/rouge/.

[2]. Alguliev, R., Aliguliyev, R., 2007. Experimental investigating the F-measure as similarity measure for automatic text summarization. Appl. Comput. Math. 6 (2), 278–287.

[3]. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., Kochut, K., 2017. Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.

[4]. Chen, Y.C., Bansal, M., 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. arXiv preprint arXiv:1805.11080.

[5]. Conneau, A., Kiela, D., Schwenk, H., Barrault, L., Bordes, A., 2017. Supervised learning of universal sentence representations from natural language inference data. arXiv preprint arXiv:1705.02364.

[6]. General purpose unsupervised sentence representations. Retrieved from https:// github.com/epfml/sent2vec.

[7]. Infersent Sentence Embedding. Retrieved from https://github.com/facebook research/InferSent

[8]. Kanapala, A., Pal, S., Pamula, R., 2019. Text summarization from legal documents: a survey. Artif. Intell. Rev. 51 (3), 371–402.

[9]. Kasture, N.R., Yargal, N., Singh, N.N., Kulkarni, N., Mathur, V., 2014. A survey on methods of abstractive text summarization. Int. J. Res. Merg. Sci. Technol 1 (6), 53–57.

[10]. Kavila, S.D., Puli, V., Raju, G.P., Bandaru, R., 2013. An automatic legal document summarization and search using hybrid system. In: Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA). Springer, Berlin, Heidelberg, pp. 229–236.

[11]. Keras: The Python Deep Learning library. Retrieved from https://keras.io/.Kim, M.Y., Xu, Y., Goebel, R., 2012. Summarization of legal texts with high cohesion and automatic compression rate. In: JSAI International Symposium on Artificial Intelligence. Springer, Berlin, Heidelberg, pp. 190–204.

[12]. Lin, C.Y., 2004. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out. pp. 74–81.

[13]. Tom M. Mitchell, Herbert Jones, “Machine learning”, Indian edition. Tata Mc Graw Hill Publications, 1 July 2017.

[14]. SukritiVerma,VagishaNidhi,"ExtractiveSummarizationusingDeepLearning",arXiv:1708.00439v2[cs .CL], 9 Jan 2019.

[15]. Bhumika, V. Srikanth,” Abstractive Summarization with NLP Makes Easier to Memorize”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-5, Jan 2020.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Data Privacy in Artificial Intelligence

Ms. Mounica B[1], Dakshata N. Ramteke[2]

[1]Sr. Asst. Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Information, data privacy and security concerns are a persistent trend that we've been reporting on nearly every year since computers started booting up. As artificial intelligence is emerging with dynamic changes ,it works with both pro-cons in field of data privacy . Artificial intelligence magnifies the ability to use personal information in ways that can intrude on privacy interests by raising analysis of personal information to new levels of power and speed while artificial intelligence also helps in analyzing and detecting patterns in large volume of data called as big data which can be beneficial for some business as well as threat to data privacy.

Keywords: Data Privacy, Artificial Intelligence, Big Data

I. INTRODUCTION

Recent news about data leaks, the lack of control over content, and significant influence of social networks has provided an increasing awareness of how social media platforms misuse personal data, which in turn has had an effect on the level of trust users have in such platforms and digital services . Many social media platforms get their (economic) value from capturing user's behaviour either directly (via services offered) or indirectly (by tracking users' online activities). With the migration from laptop- or PC-based browsing via web browsers, to consuming media on mobile devices and via dedicated apps, it has become possible to collect far more different types of data surrounding this behaviour in a far more targeted manner; now it is possible even in near real time . Combining places where people go digitally and physically offers many possibilities, but also brings about many new privacy risks. Although location data are explicitly categorized as personal data in the General Data Protection Regulation, it is not always clear what kinds of risks such kind of data poses, specifically in combination with other types of personal and non-personal data. Discussions on what personal data exactly means and how to apply personal data protection. In the context of large-scale data analytics are even more pressurizing the current landscape of data protection regulation . Slowly but definitely, companies and governments deploying big data analytics and process personal data are applying (and complying to) the GDPR[1]. Beyond the growing awareness of the need to comply. there is a wider societal need for trust in digital environments.

As we know one of the most important reasons business, especially consumer facing business, wants to have lots of data is to know as much about the market, us, as possible. Artificial intelligence (AI) has

made that focus on customers more and more easier and accurate. While business has been becoming more invasive, government are focusing more into this and passing regulations that begin to provide certain limits. Privacy matters to the all, and smart business looks at how to use data to find out information while remaining in compliance with regulatory rules.

Information, data privacy and security concerns are a persistent trend that we've been reporting on nearly every year since computers started rapidly increasing. And now, the economic stakes, social consequences and technology get more serious and complex[2]. Privacy issues used to be centered around evading online activity trackers as they follow you around with ads for things you don't want or want. Now exposed as centre to many political and ethical scandals, data privacy has become one of the defining social and cultural issues of our era. Our need to control what we hide and what we shares extends from our very person and lead to our homes, businesses, communities, and governments. And because of the pervasive nature of technology, the data it creates and carries has burrowed into our lives in ways that we now take for granted.

II. LITERATURE SURVEY

Privacy and AI are terms that are emotive and misused. AI is associated with reduce the cost of prediction. The outcome on privacy is ever more platforms of data will be utilized to predict thousands of outcomes. However, this paper will compare, contrast and critique contextualized examples to highlight the Privacy and artificial intelligence [3]. The privacy is that most of the information created and produced in our digital universe, is created and consumed by the consumers themselves: social media networks, digital TV, sending and receiving mobile images between devices connected to the internet. However it is companies that that have 80% of the responsibility and therefore the liability of privacy, copyright and compliance. The biggest media company in the world creates zero content – Facebook, the individual users create it for them and continue to use the platform despite Facebook recently paying a $5 billion dollar fine for the infamous Cambridge Analytical data breaches, as documented in the article 'How Cambridge Analytical used your Facebook data to help the Donald Trump campaign in the 2016 election' . Psychometrics, based on the five personality traits known as OCEAN: openness, conscientiousness, extroversion, agreeableness, neuroticism can assess what sort of person we are and predict how we will behave. The issue however, was always the data collection – this all changed with the combination of the internet, the emergence of Facebook, Michal Kosinski and Cambridge Analytica [4]. The predictive model that Kosinski and his team had created by 2012 could prove that on an average of sixty eight Facebook likes by the user that they could predict: the colour of your skin (95% accuracy), sexual orientation (88% accuracy), affiliation to Republican or Democrat Party (85% accuracy) [5]. What confirmed how robust and reliable the modelling was becoming, was the predictive evaluation of the personality profile. The model was becoming consistently more accurate with less information: in his findings the (Kosinski et al) [6] model was now able profile an individual in ten facebook likes, to a greater accuracy than a co-worker, Seventy likes would profile the individual better than a friend, One hundred and fifty likes better than an individual's parents.

Images and videos that are created using deep learning and contain a real person acting or saying things they didn't do or say are called Deepfakes. If used for entertainment purposes, Deepfakes are fun, but people are creating Deepfakes for fake news and knowledge and, worse, Deepfake porn. In 2019, an application called DeepNude was

launched where you'll upload any image of girls , and it might generate a real-like nude image from it. It is quite disturbing that anyone could exploit woman images available online by creating nude images from DeepNude. After too much controversy, it was shut down. Still, it's just a matter of your time that somebody can again misuse Deepfake technologies by taking your publicly available videos or photos.

In recent years, China has received severe criticism thanks to mass surveillance of its people without their consent. They use over 200 million surveillance cameras and facial recognition to keep a constant watch on their people. China also mines their behavioral data captured on the cameras.To make it worse, China implemented a social credit system to rate the trustworthiness of its citizens and give them ratings accordingly based on their surveillance. People with high credit get more benefits and low credits loose benefits. But the more severe part is that each one this is often being determined by AI-based surveillance without people's knowledge and consent.

III. PROPOSED METHODOLOGY

Every time you attend the web for searches, browsing websites, or once you use mobile apps, you are doing not even realize that you simply are making a gift of your data either explicitly or without your knowledge. And most of the time you permit these companies to gather and process your data legally since you'd have clicked "I agree" button of terms and conditions of using their services. Apart from your information that you explicitly submit to your websites like Name, Age, Emails, Contacts, Videos, or Photo uploads, you also allow them to collect your browsing behavior, clicks, likes and dislikes. Reputed companies like Google and Facebook use this data to enhance their services and don't sell this to anyone. Still, there are instances where third party companies have scrapped sensitive user data by loopholes or data breaches. In fact, the only intention of the many companies is to gather user data by luring them into using their online services, and that they sell this data for vast amounts of money to third parties. The situation has worsened with the surge in malicious mobile apps whose primary purpose is to gather even that data from the phone that it didn't seek permission. These are primarily data collection apps disguised as a game or entertainment app. In today's world, smartphones contain very sensitive data like personal images, videos, GPS location, call history, messages, etc., and that we don't even know that our data is getting stolen by these mobile apps. Every now and then, such malware apps are removed from Play Store and Apple Store but not before they have already been download millions of times. We're seeing a social shift in the long term effects of privacy as billions more in venture investing targets our personal data for resale in a multitude of ways, people are starting to more deeply question their growing lack of data privacy and control. The challenges of protection of personal data in the context of Big Data Analytics (BDA) mainly connect to concepts such as profiling and prediction supported large datasets of private data. A secondary result of big data analytics is that combinations of non-personal data can still lead to the identification of persons and/or other sensitive .A dilemma put forward by data science is that data protection and data-driven innovation is diverging,and have even opposite premises.. We need to look for new ways to guarantee the protection of personal data while retaining the potential benefits of big data analytics

The Healthcare Industry already has many protections in situ, with HIPAA, and

frequently reminds us of how they're protecting us. Meanwhile our wearables are collecting a lot of health-related data on us. Who owns that data? And now that Google has bought Fitbit, what's that getting to mean for privacy? Add the many folks that have given away their DNA to seek out ancestors and therefore the Google Nightingale project to the privacy issue and it's clear HIPAA's getting to need an update[7].In Facial recognition what feels like a prequel to Minority Report, people's physical safety and movements are at risk. Citizens are taking measures to protect themselves from detection, trying to avoid arrests at protests or simply not wanting to share their whereabouts in public settings like an airport.As reported by CNN on varied defensive measures people have taken to protect their privacy - from rudimentary scarves and goggles to incredibly lifelike paper masks used to anonymize protesters. Big Think reported on designers using LED-equipped visors and transparent masks to guard identity. San Francisco became the primary city to ban the utilization of face recognition software by the town .We now know that voice-activated devices are listening all the time. Are our phones eavesdropping on us, too, as reported on Marketplace. It's inconclusive, but the investigation suggests there could also be more ways we divulge information than we realize. We also got to mention all the Ring and Nest doorbell systems. capturing video not just from your front entrance but all round the neighborhood, as well. Ring is actively collaborating with local enforcement - a practice that's raising privacy questions at the local level. We are being swiped right, All the clicks we've left behind are being used to rate us. And that can sometimes work against us. A recent piece from The NY Times outlined the industry-for-hire that makes a score for every one among us and sells it to businesses. Right, the list goes on and this is often just the start . Data privacy concerns reach voting and what data protection means to democracy. Facebook, Twitter, YouTube, TikTock, Google all have integrated with brands to hyper target us down to the tap, touch, and like. The truth is that there is only so much regular citizens can do without laws and policies that empower citizens to retake some personal data power.

The EU's GDPR was a blunt first instrument .Just trying to turn things off by playing whack-a-mole won't work. we need new innovations focused on protections that are more conversation driven and transparent. Tech companies today are built by a number of the neatest people in business – they ought to be ready to work within the bounds of latest laws to repair this. Finding ways to claw back and respectfully manage that data will prove essential to all users. Until now, consumers have been willing to lend their data (or have unknowingly given it away) to urge convenience or information reciprocally[8] . Once they fully realize the results of this bargain they're going to be looking to government and business to safeguard data and hand control back to them, the customer. Business needs to start thinking now about how to counteract the fear and distrust flooding the marketplace. The should provide verifiable solutions, traceability and transparency. They should be willing balance upholding privacy concerns without annoying users with privacy notifications and too many restrictions.

In Europe, the General Data Protection Regulation (GDPR) that came into force in 2018 regulates the collection and use of personal data.[1] Data protection law does not refer explicitly to AI or Machine Learning but there is a significant focus on large-scale automated processing of personal data and automated decision-making. This means that where AI uses personal data it falls within the

scope of the regulation and GDPR principles apply. This can be through the use of personal data to train, test or deploy an AI system. Failure to comply with the GDPR may result in enormous penalties for the companies involved

Cape Privacy may be a company that states its focus as providing software which integrates into a company's existing data science and machine learning infrastructure, enabling all parties to figure together on projects and policies. That is true start up speak, with an initial specialise in the implementation of privacy policies at the event level. The company works to help programmers work to oversee privacy policies in the data and between applications. What its link to the legal and compliance groups is more rudimentary, there's clearly a vision showing intent to enhance collaboration because the solution grows.

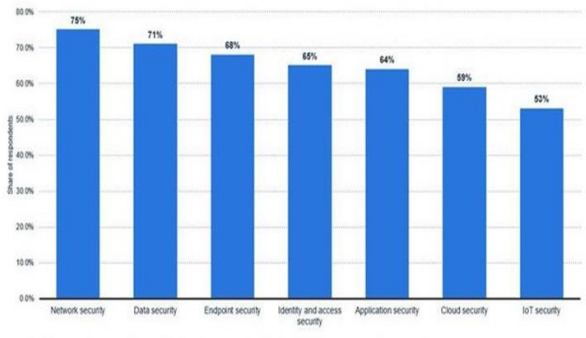


Figure 1. top AI use cases for cyber security in organization in selected countries as per past years

Now many countries have created their own data regulation policies to bring more transparency between these online platforms and the users. Most of those policies are centered around giving users more authority in what data they will share and be told about how the platform would process their data. A very documented example of this is often the GDPR law that came into existence a few of years back for the EU countries. It gives EU people more control of their personal data and the way it's processed by the businesses .Large companies like Google, Facebook, Amazon, Twitter, YouTube, Instagram, and LinkedIn literally own a majority of users' social data across the planet . Being, reputed giants, they ought to be extra careful to not leak any data to malicious people either intentionally or unintentionally.

The people of AI communities, especially the thought leaders, should raise their voices against the unethical use of AI on the private data of the users without their knowledge. And also they ought to educate the planet that this practice can cause such a disastrous social impact. Already many institutes are teaching AI ethics as a topic and also offering it as course.Finally, we should always remember that, despite government regulations, these are just policies and therefore the responsibility lies with us as individuals. We have to take care about what data we are uploading on social platforms and mobile apps and always inspect what permissions we are giving them to access and process our data, let us not merely "accept" anything blindly in "terms and conditions" that comes our way on these online platforms.

There are many concerns round the ethics in AI within the synthetic Intelligence Community thanks to the social biases and therefore the prejudices it can create. But processing personal data with AI without people's consent and further misusing it raises the concerns of AI ethics to subsequent level. AI emergence within the world remains in nascent age, and that we all need to intensify to form sure that our creating AI by misusing personal information should not become a daily occurrence within the future.

IV. CONCLUSION

Emerging artificial intelligence is affecting data privacy in very different way which should be restricting as it disobeys the basic privacy of an individual. some protection techniques are developing but it is still not clear they will be restrict this with changing artificial intelligence. Protecting privacy and securing digital data will continue to be a fundamental risk issue as AI becomes more mainstream in healthcare, raising numerous legal and ethical questions. Thus, it'll be obligatory healthcare leaders, AI developers, policymakers, data scientists, and other experts to spot vulnerabilities and consider innovative and proactive strategies to deal with them.

V. REFERENCES

[1]. Artificial Intelligence and the Privacy Paradox of Opportunity, Big Data and The Digital Universe Gary Smith Amity University Dubai Dubai, United Arab Emirates ,2018

[2]. Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation Bernd Carsten Stahl | De Montfort University David Wright | Trilateral Research,2019

[3]. J. Bossmann, "Top 9 Ethical Issues in Artificial Intelligence," World Economic Forum, 21 Oct. 2016;

[4]. W. Knight, "The Dark Secret at the Heart of AI," MIT Technology Review, 11 Apr. 2017.

[5]. C. Sandvig et al., "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms," Data Discrim. Convert. Crit. Concerns Product. Inq., 2014.

[6]. M. d'Aquin et al., "Towards an 'Ethics in Design' Methodology for AI Research Projects," AAAI/ACM Conf. Artificial Intelligence, Ethics, and Society, 2018.

[7]. S. Tan et al., "Detecting Bias in Black-Box Models Using Transparent Model Distillation," Oct. 2017.

[8]. J.A. Kroll et al., "Accountable Algorithms," Univ. Pa. Rev., 2016.

[9]. O.J. Erdélyi and J. Goldsmith, "Regulating Artificial Intelligence Proposal for a Global Solution," AAAI/ACM Conference on Artificial Intelligence, Ethics and Society, 2018.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Information Security Risk Management Framework for the Cloud Computing Environment

Dr. Arvind S. Kapse[1], Madhavi Basant[2]
[1]Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India
[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Cloud services provide organizations the opportunity for on-demand network access to a shared pool of configurable cloud environment. The security risks associated with cloud environment model are dependent on factors such as sensitivity of information assets, cloud architectures, storage cost and security controls involved in cloud services. Risk management framework is one of security assessment tool to reduce the threats and vulnerabilities of cloud computing environment. Based on risk management framework a prototype was developed using machine learning to automatically analyse the risks involved in cloud environment. [3] When used appropriately with the necessary precaution and control in place, cloud computing could yield a multitude of benefits, some unheard of until now and some yet to be discovered.

Keywords: Cloud computing, Risk assessment, Security, Data Privacy

I. INTRODUCTION

Cloud computing industry continues to grow at a fast-pace and have become one of the top technology trends. Companies can get advantage of the lower costs, efficiency and high scalability of cloud computing services. There is no upfront cost for installing and managing the software and hardware infrastructure for cloud environment [1]. Cloud computing also has brought many risks to organizations because they outsource IT resources, which results in providing access to third party. Therefore, such organizations might lose control over security of the data resource in the their cloud environment

Before moving to a cloud-computing environment, companies need to understand the business and legal risks, which is associated with cloud services, and have some effective planning in place to resolve those risks. [1] They lack experience incorporating cloud services from both a contracting and a technology domain. Companies need to review the risks of cloud environment and mitigate it. Organizations might mitigate cloud-computing risks through negotiation of contract clauses and service level agreements (SLA) and enforcement of the same [2]. The SLA should be the mutually agreed between the cloud service provider and the organization. If organizations already have agreements with cloud service provider

then they should apply security protocols in future cloud-contract negotiation.

Cloud computing environment is a major source of new vulnerabilities in therefore, it is important to establish several protocols, which will mitigate the risks, provide security and increase trust and confidence in cloud services. Risk management is one of the cloud services security mechanisms, which provides access and manage risks related to cloud computing environment and to prevent those risks from affecting organization's business goals [2].

1.1 Types of Risk in the Cloud

Policy and Organizational Risks — The risk of vendor lock-in through the use of proprietary interfaces, organizations face issues such as complying with various regulations, even though applications and data resource have been migrated from a traditional data center to a cloud service environment.

Cloud Service Failure — Cloud vendor also must deal with the risk of a cloud service failure and be prepared to mitigate these risks (eg: disaster recovery and business continuity planning). There is another type of risk to consider is, the loss of business reputation of an enterprises. For example, a cloud security breach could result in damage to the reputation of enterprises using that cloud services.

Technical Risks — Enterprises that uses cloud services also face technical risks such as resource exhaustion, which can occur when not managing cloud services resources effectively. Resource exhaustion can cause degradation of services and, in some cases, a failure to provide services at all in the cloud environment.

Interception of Data — The interception of "data in transit" to or from a cloud environment. This can be resolved by ensuring that all data transmissions are encrypted and that the endpoints for any data transmission are authenticated to ensure that they are legitimate.

Distributed Denial-of-Service Attacks —The cloud vendor and their Internet service providers are typically responsible for dealing with DDoS attacks. Organizations should discuss DDoS with cloud vendors to ensure that the proper controls are in place to resolve a DDoS attack. [8]

Legal Risks — Enterprises should maintain compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) or the Payment Card Industry Data Security Standard (PCI DSS). Many enterprise find it difficult or even impossible to achieve compliance while using cloud environment. Cloud vendor should thoroughly research any applicable regulations before planning a cloud service migration.

Changes in Jurisdiction — Enterprises may face varying requirements if cloud vendors are located in different geographic locations, each of which has its own set of law.

Data Privacy —Cloud vendor state in their contracts that they have the right to monitor customer data in the cloud environment and harvest data from it, then sell this data to third parties.

II. LITERATURE SURVEY

The literature review of this paper begins with the status of cloud contracting in Canada and the US government, followed by a review of the rewards and risks of cloud computing. It identifies risk management strategies through negotiation of key contract terms and SLAs [7]. The main goal of the research was to identify major contract terms needed

to mitigate risks as organizations move to a cloud-computing environment.

The risk management framework provides key terms to organizations, a checklist of cloud services clauses to add in the contract in order to protect their interests from a business and legal contracting point of view. This paper also includes key points important for negotiation of cloud services contracts, which included governance and Software as a Service strategy, project management and vendor performance management. [4]

2.1 Service model of cloud computing:

Software as a Service (SaaS) – Cloud services delivers an application, which is already customized with all of the required hardware, software, operating system, and network to be accessible by various consumers by using the internet services without the need to install software on the servers.

Platform as a Service (PaaS) – Cloud services offers an environment and all the requirements such as software tools, libraries, software programming languages, and services for cloud consumers to develop or install their own software applications and tools). The applications are then delivered to the users via the Internet services.

Infrastructure as a Service (IaaS) – Cloud services offers fundamental computing resources such as processing, servers, data storage, and networks for cloud customers to install and run their own operating systems and applications. Cloud computing services can be deployed in four ways dependent on consumers' requirements such as public cloud, private cloud, community cloud and hybrid cloud.[4]

2.2 Cloud services deployment models

Public Cloud – Cloud environment are made available to public customers over a public network. Multiple enterprises can work and access the provided infrastructure and resource at the same time. The public cloud model is managed and owned by a third party: a cloud vendor.

Private Cloud – Cloud services are dedicated only to a specific consumers (enterprises) and offer the security and control over client data resource. A private cloud may be controlled and owned by the organization itself or a cloud vendor

Community Cloud – Cloud infrastructure is equipped for a group of consumers (enterprises) which have the same shared requirements. They can share resources by using the connections between the associated enterprises. The community cloud can be controlled and owned either by the relevant community enterprise or a cloud vendor.

Hybrid Cloud – A hybrid cloud is a mixture of two or more types of cloud deployment models such as public, private, or community. They are connected together to allow for the transfer of data resource and application between them but without affecting each other services and structure.[4]

Risks can be reduced during various stages from selection of the cloud vendor, pre signing of the contract and post signing of the contract. Below are the mentioned stages for risk management mitigation :
Due Diligence Stage (cloud vendor selection)

Business, Legal and Regulatory Risk Mitigation Stage (negotiation of contract)

Vendor Performance Management(on boarding, during and post contract terms, review of SLA metrics, audits).

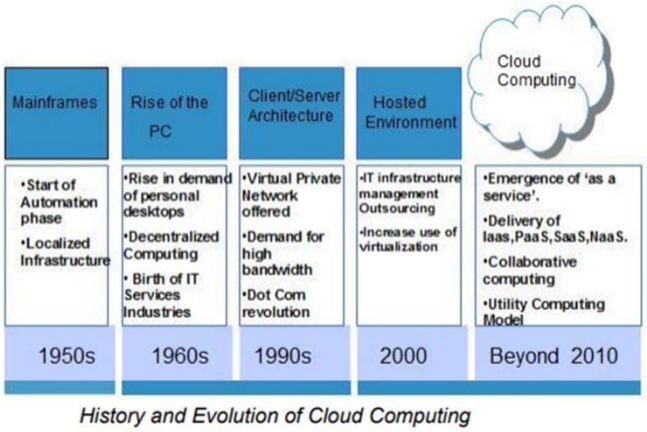


Figure 1: History and evolution of Cloud Computing

III. PROPOSED METHODOLOGY

Many organizations have switched to cloud computing environment, but often this adoption has not involved full-fledged risk management methodologies. Instead, they trust cloud providers to take care of security. Security is an important consideration for those looking to access cloud technologies. Cloud Vendor must also be concerned about privacy, availability and reliability, disaster recovery, business continuity, scalability, compatibility and standardization.[3]

Risk management of cloud computing is a mechanism for managing the risks or threats facing by an organization that may result in damaging the resources and to provide best course of action for identifying and resolving risk issues. Some of the objectives of cloud computing risk management are increasing system security, protecting and enhancing the organization's assets, optimizing operational efficiency and making well-informed decisions

The key points for organizations to mitigate the risks associated with contracting with cloud services is mentioned below:

1. Organizations should perform due diligence to reduce the risks at the pre-contract level
2. Cloud services contract and SLA metrics should be incorporated into cloud vendor contracts to mitigate risk in cloud environment.
3. Cloud service providers are becoming more willing to accept some of the risks and are starting to work with organizations to negotiate mutually acceptable contract terms.

3.1 Effective Risk Mitigation Steps [5]

STEP 1: Examine the business context- This refers to assessing the types of services and data resource handled (eg: customer records, credit card numbers, and prescriptions). Cloud vendors face numerous negative scenarios when analyzing these threats for each service and data resource.

STEP 2: Review application security - Assessing the security of applications is a critical step in risk management. It is important to determine what features an application offers to mitigate risk. Cloud vendor also should consider how robust these features are. Suppose if an application encrypts the data at rest to protect its confidentiality, what encryption algorithm and key length are used, and how strong are the algorithm and key combination.

STEP 3: Construct a data-centric governance plan – Data governance involves the management of data resource throughout an enterprise, such as making backups and archiving old data resource. A survey which was conducted in 2013, shows that nearly half of all enterprises didn't have a data governance policy. This can increase the chances that data resource will be misplaced or compromise. It is important for organizations who are using cloud services to know where their data resource is stored and to restrict the access. The virtual machines on which the data resource resides should be authorized to restricted individuals only.

STEP 4: Test and retest. After a data-centric governance plan has been implemented, cloud vendor should periodically test data resource. This includes actions such as verifying the integrity of backup data resource, ensuring that archived data resource is stored securely and making sure that unnecessary data resource is destroyed properly. Authorized users must be able to find data on demand, such as that needed to respond to e-discovery requests. Testing should be performed periodically on data resource to ensure that these controls works as intended.[6]

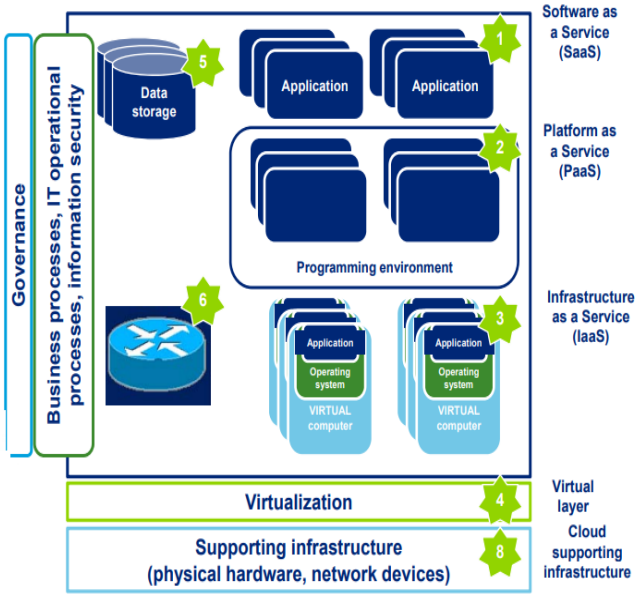


Figure 2: Cloud Computing Framework

IV. CONCLUSION

Risk management framework can be decided based on its effectiveness on a cloud environment. Applying cloud computing risk management model without the proper care, due diligence and control, is bound to cause major unseen problems in future. By being aware of the risk and other issues related to cloud computing environment, organizations are more likely to achieve their objective as they manage the risks in their dynamic and cost effective environment that likely will become the most secure computing model of the future. Cloud computing is relatively new in its current form, given that, it is best applied to specific low to medium risk business areas.

V. REFERENCES

[1]. Kathy English, "Cloud Computing: Risk Mitigation Strategies from a Contracting Perspective" April 2015

[2]. Vinit Kumar, "Information Management in Cloud Environment: Risks and Mitigation Strategies" February 2018

[3]. Erdal Cayirci, Alexandr Garaga,"A risk assessment model for selecting cloud service providers" September 2016

[4]. Mariana Carroll, "Secure cloud computing: Benefits, risks and controls" September 2011

[5]. Chiang Ku Fan,"The Risk Management Strategy of Applying Cloud Computing" 2012

[6]. Rana Alosaimi,"Risk Management Framework for Cloud Computing : A Critical Review", August 2016

[7]. Jordi Guitart, "Introducing Risk Management into Cloud Computing", August 2016

[8]. Usha Kiran Marichetty, "The Use of Effective Risk Management in Cloud Computing Projects", October 2017

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Automatic Waste Segregation and Management

Dr. Avinash S. Kapse[1], Mr. Gaurav Prashant Kale [2], Mr. Kishan Vijay Nardiya[2], Mr. Praful Devendra Masne[2], Mr. Shubham Kishore Dukare[2]

[1]HOD, Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

With the increase in the development of smart cities, the idea of keeping cities clean is the highest demand. The amount of waste produced is too large and the manual effort required to treat it is very tedious. With the evolution of technology in all fields, automated means can be adopted to prevent stacking of garbage. The waste sorter is designed to facilitate waste disposal collection. The system consists of three trays, each for wet, metal and dry waste. The conveyor belt system detects the incoming waste and classifies it as metal, dry or wet using various sensors connected to the system and divert it into the respective bin. The deviation procedure is performed by the servomotors which are programmed according to job. This makes it easier to deal with different types of waste as per requirement. The level of waste in each of the waste bins is monitored using ultrasonic sensors present in each trash can. The notification is then sent to the authorities concerned to empty the bin. The entire configuration results in automation and thus reduce human intervention necessary to sort waste and allows for success timely collection of garbage in the bin. The system is driven by a microcontroller

Keywords:- Arduino UNO, and the sensors are programmed using the Embedded language vs.

I. INTRODUCTION

With increase in population from year to year, the amount of waste generated increases considerably. This has led to many dangerous problems. The accumulation of waste in large areas of land results in the formation of hazardous waste consequences. The smell of rotten waste pollutes the environment by emitting a foul odor. The disposal of waste in water bodies contaminates all connecting oceans and seas that affect the quality of the drink water and also the life of aquatic animals. The toxic gases are released into the air and in turn the whole the ecosystem is affected. Waste management is therefore a very serious problem in our time. If the waste produced is effectively managed at the source level, a lot can be changed and prevented.

Separation of waste into wet, dry and metallic categories can help dispose of waste appropriately and in the implementation of the principle of reuse, reduction and to recycle. Wet waste can be broken down to produce manure for plants, metal waste and dry waste can be recycled. Thus, the automatic waste separator has many applications in waste management. The system separates the waste into 3 different bins under the wet, dry and metallic category. Different sensors are used for detection of the type of waste. The level of the garbage in the bins

is constantly monitored so that the bins do not overflow and they are emptied on time. The notification is sent to the relevant authority with the where the bin is placed.

II. RELATED WORKS

Waste Segregation using Deep Learning Algorithm

This paper presents an automatic classification of waste based on Convolutional neural network. It separates the waste into non-biodegradable and biodegradable categories using Image classification based on deep learning. The notion of Deep Learning allows "the processing of many layers through computer models in order to learn data representations with abstraction of several layers.
This is suitable for a huge amount of waste. Classification of the real-time materials are done using a webcam with python index package. Tensor Flow and spyder are open source software libraries used. The training process is very time consuming. It prevents manual labor at a large Degree. The system reduces the pollution level to a large extent and has the opportunity to become an important framework in the years to come.

Waste Management using Internet of Things (IoT)

This paper examines the amount of time and money that is wasted due to scheduled visits to trash cans in the area. There are times when workers find the bins are overflow and other times when the trash cans are not even to the brim. This causes one of the main reasons for inefficiency in the proper collection of waste and is also harmful to the environment.
Hence a smart trash can that solar powered and compact waste is developed. He has sensors present in it which filter the amount of waste has been collected and therefore compacts the waste so that it can hold up to 10 times the ordinary containers.
He can remotely transmit the fill level using the cloud server. Tray can be used as a Wi-Fi hotspot and is powered either by battery or by solar energy. It detects the amount of waste is inside the compartment and remotely transmits the fill level data to the cloud server. The result is a productive use of time and makes the roads more hygienic.

IOT Based Smart Garbage Monitoring and Alert System Using Arduino UNO

It sets up a waste disposal system by notifying an alarm to the municipal corporation's web server. He notifies the time of cleaning the waste in the trash by maintain legitimate control over the level of waste filled. The ultrasonic sensor connected to the Arduino UNO is used for measure the waste collected in the container. All the procedure is maintained by incorporating and using GSM and GPS based on the Internet of Things. An android the application is created and it is constantly connected to the municipal web server to download the alert received from Arduino and provides remote observation.

Standalone Frequency Based Automated Trash Bin and Segregator of Plastic Bottles and Tin Cans

The article shows how the piezoelectric amplifier the frame can be used to feed the input signal and noise can be eliminated using a comparator. The average the frequency response of the object triggers the frame as it hits the platform. The Arduino decides what to do next processing steps. Galvanized iron is used for the manufacture of the platform and it is associated with a DC motor which isolates the elements. There is an overlap in the frequencies of metal and plastic pots found with the middle frequency values despite the fact that there were hardly any example causing an overlap. When the first hit the bottom of the plastic container on the platform the results demonstrated the accuracy of the

frame when the the greatest success of the scene was initially more contrasted with the lower value.[4]

Eco-Friendly IOT Based Waste Segregation and Management

Viable and effective techniques for assortment and isolation of waste at the residential level according to the nature of the composition are concentrated in this article. The waste is properly stored in its own parts of the bin made of metal, plastic and biodegradable. Biodegradable waste is distinguished by rejection methane gas and after a certain amount of gas, liquid is sprayed on it. The filling state of the bin is observed and the data is sent using the Wifi module. At the moment when a certain level is reached, a notification is sent to the client to advise that the specific portion is covered and that it must be discarded. The entire configuration of the tank is controlled by STM32 microcontroller. Likewise, both plastic and metal waste are recognized and isolated in their separate areas using capacitive and inductive sensors respectively.[5]

Automated Waste Segregator

The document describes an easy-to-use and inexpensive way to segregation by introducing automated waste Segregation. It separates waste into wet, dry waste waste and scrap metal. It uses capacitive sensors for perceive wet and dry waste and parallel resonance impedance identification component for perceiving metal things. With an orderly priority to separate the wet, dry and metal waste, only one type of waste can be isolated at once. Mixed waste is isolated by improvements that use buffer spaces. Metal detection the waste is comparatively less, and the whole module can be put into a solitary stage where the entity is stable to ensure better results.[6]

A Novel Approach For Waste Segregation At Source Level For Effective Generation Of Electricity –Green Bin

The paper describes the innovation of automating the insulation of dry waste in plastics, metal, paper and glass at an accessible commercial level. Isolation of waste is determined and influenced when the waste becomes contaminated. The green bin is intended for mechanized isolation of waste at the local level. It includes inductive metal sensor, capacitive humidity sensor, odor methane sensor and sensor. Capacitive based humidity. The sensor is used to decide the level of humidity present in the waste. Inductive based metal sensor is used to detect metal content of the waste. The claim of food waste isdetermined by methane sensor. [7]

Methodology

The automatic waste separation system is driven by the Arduino UNO microcontroller. All components that are connected to Arduino UNO are programmed using the Arduino IDE. The program is written in embedded C language and it reads the input / output pins of the Components. The conveyor belt system moves when smells like garbage. The servomotors are present to deflect wet, dry and metallic waste in specific bins. The metal waste is detected by inductive proximity sensors. Dry waste contains paper and plastic which are differentiated using the capacitive proximity sensor. The wet waste is examined using the humidity sensor. The level measurement of the bin is calculated by the Ultrasonic sensor connected to the edge of the bin. When the trash can is full, a "BIN IS FULL" message is displayed. sent to cleaning authorities. The message is sent using the GSM module which ensures communication between the bac and the authority. The location of the bin may also be sent. The location is known with the use of GPS module connected to

the system. Fig. 1 shows the step-by- step operation of the entire system.

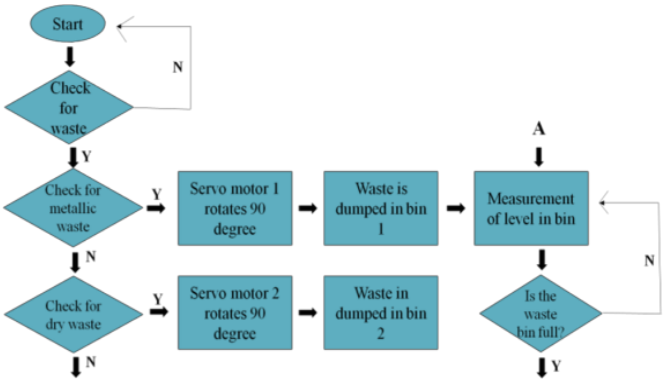


Fig. 1 Data Flow Diagram

III. MATERIAL REQUIREMENTS

ARDUINO UNO- It is an "open-source electronics platform "dependent on simple hardware to interface and execute programming. The personalization of microcontrollers is made using Embedded C and C ++ programming codes. The Arduino microcontroller gives a Integrated Development Environment (IDE) that supports different programming languages.

1. FOR SEGREGATION:

CONVEYOR BELT which ensure the circular movement of the rotating belt about them. The pulley which is powered is the drive pulley and the non-powered pulley is the idle pulley.

DC MOTOR - It is used to drive the conveyor belts. The speed is also affected by the change in the quality of the current in the field windings.

SERVOMOTOR- It is used to divert waste to the respective bins. A servomotor is defined by "a rotary motor actuator or linear actuator which takes into account Digital or analog input

HUMIDITY SENSOR - It is used to identify if the the garbage is wet or dry. The moisture content of the waste is tested and, therefore, it is deposited in the trash can.

PROXIMITY SENSOR - The presence of objects is detected without any physical contact using proximity sensor. The Inductive proximity sensor is used to identify metal waste. For identification of paper and plastic Capacitive A proximity sensor is used. It also differentiates between them like paper and plastic have different permittivity value.

2. FOR THE DETECTION OF THE WASTE LEVEL:

ULTRASONIC SENSOR – It is used to monitor the level of garbage in the bin. The ultrasonic acoustic sensor is divided into three categories: receivers, transceivers and transmitters. The transmitters radiate the ultrasound by converting electrical signals into ultrasound. It is then reflected by the obstacle and received by the receiver which converts ultrasound into an electrical signal. The reflected signals are used to interpret the position of garbage in Trash.

GSM MODULE – Mobile phone technology is used in It's like a sim which is connected to the Arduino with various types of data obtained from the board. The board has pins to take + 5V or other power and mass values Connections. The TTL-level serial interface with the host is implemented in this technology.

GPS MODULE- with the Arduino UNO which uses global positioning. The the registered location is sent to the authorities using GSM module integrated in the unit.

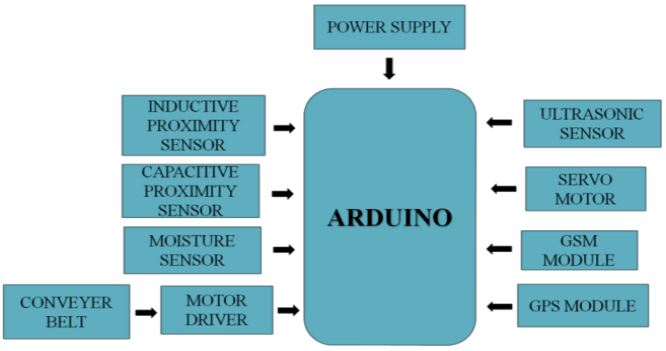


Fig. 2. Block diagram of the proposed system

IV. SOFTWARE REQUIREMENTS ARDUINO IDE-

A multiplatform application including functions that are encoded in Embedded C and C ++. The programs are written and downloaded to Arduino boards using the IDE. In this system, the program is written in C on-board for the operation of the hardware components. The program included separate methods for detection of metallic waste, dry and wet. One method was written to send the message to the authorities along with the location of the bin.

V. CONCLUSION

There is a production of 62 million tonnes of municipal solid waste (MSW) every year in urban India. of which 70% are collected and 20% are processed. Because of the nature of the different types of waste, it is very difficult to to throw out the trash. Segregation plays a very important role by reducing waste by reusing. Treat waste too becomes easier if segregation is done at the grassroots level. The fig. 3 shows the increase in waste generated by the year 2016, to 2030 until 2050 in various parts of the world.

This therefore indicates the increase in the need for efficient waste treatment to maintain the environment balanced. The model developed in this article is efficient and durable because it requires less energy for its operation and no human surveillance. The model can also detect when the tray is full asking the authorities to come and pick it up. This effectively reduces labor, waste of time and fuel required by the collection van. This model fits perfectly as a replacement to older bins and works well with the idea of smart city.

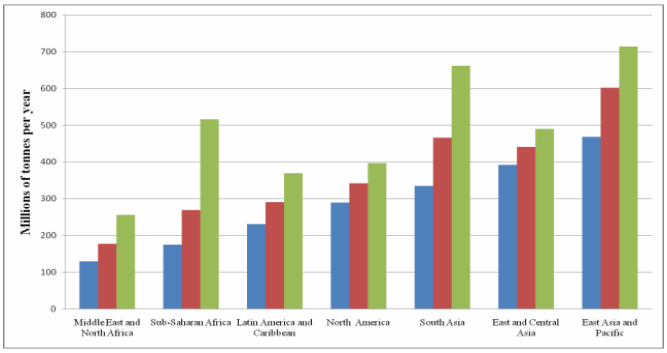


Fig. 3. Estimated increase in waste production across the world till the year 2050. The blue data represents the year 2016, red represents 2030, and green represents 2050.

VI. REFERENCES

[1]. R.S.Sandhya Devi, Vijaykumar VR , M.Muthumeena, "Waste Segregation using Deep Learning Algorithm", International Journal of Innovative Technology and Exploring Engineering

[2]. Himadri Nath Saha, Supratim Auddy, Subrata Pal, Shubham Kumar, Shivesh Pandey, Rakhee Singh, Amrendra Kumar Singh, Swarnadeep Banerjee, Debmalya Ghosh, Sanhita Saha, "Waste Management using Internet of Things (IoT)", IEEE 2017

[3]. K.Harika, Muneerunnisa, V.Rajasekhar, P.Venkateswara Rao, L.J.N SreeLakshmi, "IOT Based Smart Garbage Monitoring and Alert System Using Arduino UNO"

[4]. Marloun SEJERA, Joseph Bryan IBARRA, Anrol Sarah CANARE, Lyra ESCANO, Dianne Claudinne MAPANOO, John Phillip SUAVISO, "Standalone Frequency Based Automated Trash Bin and Segregator of Plastic Bottles and Tin Cans", IEEE 2016

[5]. Santhosh Kumar B R, Varalakshmi N, Soundarya S Lokeshwari, Rohit K, Manjunath, Sahana D N, "Eco- Friendly IOT Based Waste Segregation and Management", IEEE 2017 Rajkamal R, Anitha V, Gomathi Nayaki P, Ramya K, Kayalvizhi E, "A Novel Approach For Waste Segregation At Source Level For Effective Generation Of Electricity - GREENBIN"

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

A Big Data and the Machine Learning Algorithms

Prof. Pradip Sitaram Ingle[1], Mr. Vaibhav Subhash Borade[2], Mr. Akshay Shivaji Sonune[2]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

The Mayor of central banks discusses the dental institute of big data processing on big topics. Microdots are used with automated redesign applications in a variety of cases, including investigation, monetary policy and the established financier. Central banks are informed that they use big data for supervision and regulation (suptech and regtech applications). The quality, sample, and presentation of the data are important data for the central banks, even though the legal security on the lathe is due to the privacy and confidentiality of the data. Information institutes provide a variety of information that can be assessed on the basis of the infrastructure to be assessed, the infrastructure, to the humanitarian need. Cooperation between public authorities allows the capacity of central banks to be recovered, sorted, and analyzed macrodactyly.

I. INTRODUCTION

The world is changing and so is the way it's measured. For decades, policy maker assess the state of the economy. Collecting this data requires substantial effort and publication often occurs with a delay of several months or even years. However the The last few years have seen an explosive growth in the amount of readily available data. NewData collection and dissemination models allow analysis of large amounts of data.in real time. An important factor in this development is the advent of the information age, and especially the smartphone and cloud computing: individuals and companies produce unprecedented amounts of data being stored for future use on servers of technology companies. For example, billions of Google searches every day reveal what people want to buy or where want to go for dinner. Social media post and the private sector have relied on data published by official statistical institutions to allow market participants to track the dissemination of information on social media. Companies record every step of their production or sale process, and payment transactions and e-commerce create a digital footprint. The additional catalyst in the form of creation of big data, especially financial data, has it has been the Great Financial Crisis (GFC) of 2007-09. The GFC laid bare the need for more disaggregated data: a relatively small but interconnected bank like Lehman The brothers could bring down the financial system because it was so interconnected. Regulatory and reporting requirements established after the GFC have increased the data reported to central banks and supervisory authorities, and additional work to improvement of central bank statistics is ongoing (Buch (2019)).

The advent of big data coincides with a quantum leap in the technology and software are used to analyze it:

artificial intelligence (AI) is the topic of the day and allows researchers to find significant patterns in large amounts of data. For example, natural language Processing techniques (NLP) convert unstructured text into structured data that Machine learning tools can analyze to discover hidden connections. Network analysis it can help you visualize relationships in this high-dimensional data. For the first time in history, it is possible to produce areal-time image of economic indicators such asconsumer spending, business sentiment or people's movements. These developments have spurred central banks' interest in big data. Growing The interest is reflected in the number of speeches from central banks that mention Big Data. and to do so in an increasingly positive light (Graph 1). And yet big data and machines learning pose challenges, some of them more general, some of them specific banks and supervisory authorities. This article reviews the use of big data and machine learning in the central bank community, taking advantage of a survey conducted in 2020 among members of the Irving Fischer Committee (IFC). The survey contains responses from 52 respondents across all regions and examines how central banks define and use big data, as well as what opportunities and challenges they see.

II. WHAT IS BIG DATA

Wondered how much data is generated in the format text, email providers videos searches and music approximately generated every month by a single smartphone user in this number 5 billion smartphone users that even process is in fact this is quite a lot of computing systems and this massive amount of data is termed as big data generated from minute 2.19 snap on Snapchat 3.8 million searched on Google 1 million people Login to Facebook 4.5 million videos watched and YouTube 188 million email so how are you classify any data as big data this is possible please provide us today and value for the Healthcare industry hospitals and clinics across the world for your 2314 how can I am the danger 2314 data collected in the form of creation wickets in tests of all the status generated at a very high speed it reduces the velocity in The Dictator the various data types of structured semi structured and unstructured data examples include exam bonfires and x-ray images accuracy and trustworthiness of The generated data is termed as previously stated that it is better treatment and reduced cost this is known as the plan you think they used to process this big data store and process the spectator disturb lyrics and Spark example and see how to store hours and processes big data Hadoop uses a distributed file system known as Hadoop distributed file system the storing data if you wish you try you finally broken down into smaller chunks and store it in there is machines that when you read the file you also make copies of it would do it this way you store your pic there in a distributed then make sure that even the one machine fail you did it safe on another not produce technique is used to process big data link is broken into smart b&b Now instead of one machine Re machines Dab and thanks to the processing machine you did it safe on another not produce technique is used to process in data link is broken into smart b&b now it's a one machine Re machine take up its task and completed in a fashion and the sum of reserves and thanks to this the process becomes easy and fast this is known as parallel processing not be used toward and process of big data and analysis data from numerous applications in games like he was free and interior designers news today designer understand in which state produces fast we start work with my this insight into the river comes to your mind of the game and improve the User experience which in turn reduces the customer and made Stewart and process static data we can analyze data from numerous applications in games like industry and Call of Duty designer News Today understand it which state produces all we start work with my this is I can help them overcome story Mein

became and improved User experience which in turn reduces the customer churn rate of disaster management in 2012 he was IL picture of disaster management in 2012 he was used to gain a better understanding of the Storms effect on the east coast of the US and necessary measures were taken it could protect the Hurricane from five days late when it was impossible to hear these indications of help you can be accurately process and analyzed using video actually processed and analyzed as a question which of the following statements is not correct about attitude distributed file system in HTML is the storage layer that you did I get stored in a distributed manner in HTML c first performs processing of data stored on Fast and Furious which of the following statements is not correct about Hadoop distributed file system it at a first is the storage layer and you did I get stored in a distributed manner in htr380 performs processing of data are stored on top of that and it's in a team of the given and video actors in the comment section below 3 lakh winner will receive Amazon gift voucher plan twerk big data that is what you think will be the most significant impact of big data in the future is now in the comments if you

III. HOW BIG DATA WORK? WHAT IS BIG DATA USED FOR?

Anything anybody does is going to have some predictive power for other things to do Stephen a form a data scientist in Google in his book everybody lies beta new Tata and what the internet can tell us about who we really are very fascinating study researchers use Microsoft print to look for people who search for diagnosed with pancreatic cancer and then people who never made that search Angel look at all the health symptoms suggest that they meet in the lead-up to the diagnosis on your diagnosis they found that they were very me a patterns of symptoms that were far more likely to suggest the future diagnosis of pancreatic cancer for example searching for in digestion in abdominal pain but evidence for pancreatic cancer was searching for just in digestion without abdominal pain without abdominal pain was much more and likely to have pancreatic cancer what we to understand people to know what they are interested in or out thinking it to do a survey to actually asked them different questions but we all know that people like people can't answer certain questions given to their loved ones this is where big data becomes very interesting things that people like not ask their friends family or doctors they ask Google data scientist at Google search data and the result is something that seems very different from what we generally tend to believe of humans they have used this data to understand suicide it's been found that one can predict how many suicide there will be an area based on Google searches so when more people are searching how to commit suicide or how to give yourself it turns out the day will be more to side in that area to use it to understand why people think of suicide on what do people search before they look more information on how to commit suicide gets in inside on why people are societal it is found that 40% of people before the search for suicide search for some health condition before so health seem to be a big reason more than financial problems or relationship problems in the number one health problem that causes suicidal thoughts is depression and human behavior was extremely Complex it is no surprise that Mini existing findings in the fields of psychology and medicine are conflicting the more data we collect the better understanding of human behavior will be instead of relying on theory driven methods as often the case in psychological studies that the approaches in crime country and let me know theories arise directly from at the sheer volume of data available available on the internet is meant in finding a way to utilize the available data to understand market trends in customer behavior it in efficient application of the collected data that is used to predict tiffin market in customer Trends Every

time We type into a search box will you been something about ourselves as millions of us are searching for questions on things to buy places to meet friends produce a map of our collective home centralized Data Analytics and strategies are made accordingly to provide a personalized experience for the customer but there are potential downsize to the collection and use of big data technological expertise in civil libertarians have consistently want about security issues and big data privacy risk if you use a variable like a smart watch that your heart but it to use Netflix to watch a show of these activities we with digital trail of data which is collected for analysis of premature Corporation Sachem Facebook apple and Amazon use Big Data Analytics to collect and analyze a daughter are Dada

IV. DIFFERENT TYPES OF BIG DATA

That we are going to handle use amount of data on the this big data and that's why does become a subject that we are having military category of data will be coming here what will be our structured data and unstructured data at the last one is with the semi structured. Data structure determine that will you determine which can be represented in the form of rows and columns and that is known as the structured data as an example we can say that the respective web blocks that differed in data which should be generated by the sensor and say if we do in survey on overpopulation population will be getting our structured data. Structure data is example of database in unstructured data that data will be in the form of SIM, data file, PDF ,images, videos on the images sent by the satellites and different machines and data on tractor videos cheated on the category of unstructured data that is also used type structure theatre now it is where having 20% but , unstructured data is now a days is having almost 50% in case of semi structured data to some extent it is structured to some extent it is unstructured as an example we can be discussing the XML file, JSON files and soon they will be falling under the category of semi structured data.

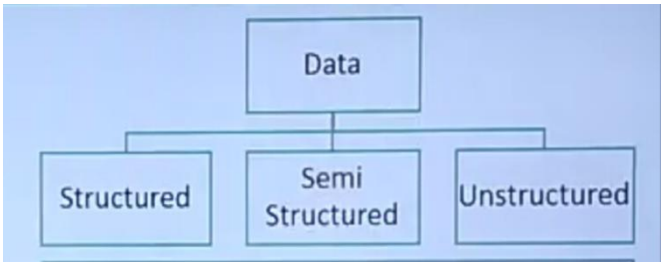


Figure 1. Types of Big Data

1. structured data
2. unstructured data and
3. semi structured data

1. Structured data:

Going for this database you can find that this is a good example of a structured data. When having certain columns so that is the employee ID , number, name, age, department when and salary and hate when having that list into rows containing the respective information so here when having the records and hearing the column structure data type of data which are stored already in an order and there are nearly 20% of the total existing data. Now a days and their data structure and all the data generated from sensors with long and deep Lock aunties all machine generated structured data is a weblog Elizabeth insert data and so that human generated structured data Atos Wiki are taken as information from in human like their names, addresses, Gender and date of birth and so on to the example of structured data.

2. Unstructured Data:

Is database concentrate is the unstructured data so, the unstructured data have no clear format in storage and we can store structured data in rows and column database path unstructured data cannot be stored like that so unstructured data cannot be stored in the form of rows and columns and when having at least 80% of the data send of the data. Now it is existing

which is feature on structure images and big data, Audio images are categorized as mission generated on structured data show the images which will be sent by the respective satellites can be treated warning in the category of unstructured data. Various types of human generated on structured data is the images ,videos ,social media, data, except for example of structured data at the text documents, PDF, images and videos and Central to this is the during on this is the concept of this on structured data.

3. Semi structured Data:

It is very difficult to categories this type of data from to structure your order from them they will be looking as on structure so that's why these data are known as semi structured data. We cannot stored this type of data using traditional database format and what it contain some organizational properties and examples of semi structured data are spread sheet files which you have in our excel in the work we are having the file specified is a good example of semi structured data. In the XML and JSON documents that are the Java Object Notation for extended markup language that is a full form of XML and NoSQL databases are the data items which are falling under this answer.

Advantages And Limitations of big data

Awareness of big data first of all we are going to discuss about that when is of big data first advantage is

- **Quick decision making**

You know big data help the business executors and managers to make better and informed business decisions. Do big data that helps in quick decision making.

- **Fraud detection**

Banks and other Financial Institutions to detect frauds like Foodland purchases with credit cards even before the cardholder gets to know about it fraud detection.

- **Improved customer service**

Big data provides massive information's regarding the customers and their tastes and preferences an analysis of such data helps the company to provide better Customer services next and when it is.

- **Competitive advantage**

Big Data Analytics helps the business to gain a competitive edge over its rivals over its rivals next one is.

- **Better sales inside**

Real time analytics tells us how this is a going so big that it is very helpful to have an idea about the sales agents.

Limitations of big data

- big data requires massive storage
- Big data may include unwanted or unstructured data
- Big data can be used for the manipulation of customer records
- Big Data Analytics may not be suitable in the short run
- Big Data Analytics may subject to technical glitches
- Big Data Analytics may subject to security threats

Problems on big data:

Problem 1:

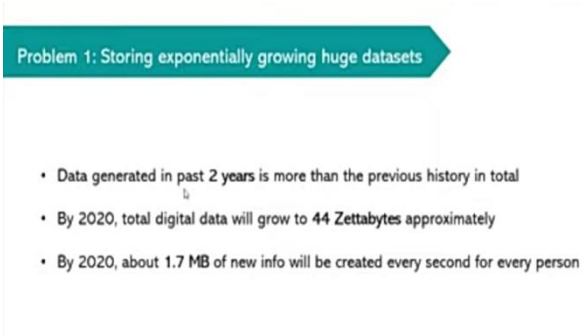


Figure 2: Problem 1 (Big Data)

The first problem that we have is storing exponentially growing to status with this already told you ok have a lot status it and for example if your

storage is limited to some 10 PB ok you cannot assume that the data will always be falling under 10 PB water one day the data is suddenly more falling under 10 PB water one day the data is certainly more on example for this the festivals feeling that we have the online sales so the website traffic is more than that more than you predict all right and sometimes you are not able to place your order as a user ,as a customer and sometimes the company will not be handle the lock lot of traffic the decade and a lot of people who are visiting the website then who are trying to find the products and in so many previous other great festivals and all they had to create online festivals there were so many issues that happened which actually spoilt the entire business of few famous Shoppers ok so we have to store the growing you status its, so it becomes a great challenge you have to make sure that the data is growing you know properly like a it certainly doesn't grow more and it is consistent so we cannot predict that but we should make sure that our system is capable enough to handle the traffic traffic to the same example that we saw in the initial volume data generated in past two years is more than the previous history in total so imagine if we are not ready to save or store all the data properly we would be the losers we would be not able to properly analyze it alright so my 2020 1.7 MB of new info will be created every second for every Passband I am sure it's more than that alright so this data story is practically difficult alright you have to make sure that you have a system I would say all rights you have to make sure that you have a system using which you can store your data which is like kind of dumping the data would pick as and when the data comes in you have to first dump it and then make sure which data you want in which state are you don't want rather than missing to store the data when you missed to store all the growing information you when you miss to store all the growing information you will be losing some really valuable data so this is the first problem storing exponentially growing huge data sets so as the picture depicts

Problem 2

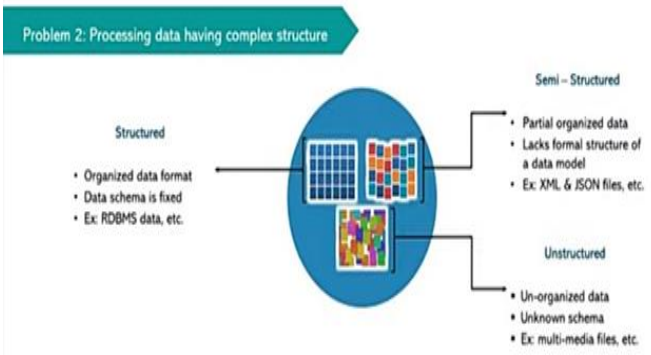


Figure 3: Problem 2 (Big Data)

Second problem is processing data having Complex structure so this was the one that we were talking about a variety when the data is of different variety we saw that it was unstructured structured and semi structured ok so when the data is having a complex structure processing the data is very difficult alright so what is it so when you take structured data and the data is in an organized format so data schema fixed ok so data cleomes fixed structure to it example is RDBMS data so few see you can identify you can identify the data based on how it is stored and based on the particular schema. Example all the HR related data would be stored under the head HR schema and FEMA and accounting details could be stored under the account scheme up so you have a prop organized data format in the structure 1 so, this is a simple thing when you when the data is in a proper structure it is, it becomes quite easy to process or simpler to process because you have it all their coming to the next to which are which are probably A B complex structure as I told you it is partially organized lacks the formal structure example XML JSON except so what we do here in the semi structured data you have to convert it to structured information and that you have to process using the structured data processing engines or your RDBMS and good example is even you know

in Oracle database you can store the data in XML format and you have the option of taking the XML and processing the using some XML queries XML query is there but again that that is possible only for unlimited data when it comes to data which grows and it is very expensive to put it in your Oracle database or any other Rachis BMS and storage and processing it. Where you can process the semi structured data, so here comes the no sequel data bases are the not only sequel data bases so no sequel are Himalayas when it is schema less what happens is you can store the data which comes in a semi structured format like cheese XML JSON so you can store the data in your NOC in your no sequel and it is it comes of different data models in structured usually it is RDBMS model or relational model or you can also call it as the table and model but here you do waste on your type of data you can store it as a document you can store it and graphs and so many other options are also available depending on the nature of your data. You can choose to use the no sequel database that you want this no sequel database is mostly open source and based on the category of data you want you can use and it tries to hide the complexity from you and helps you to process the semi structured data in a better way.

Solution on big data:

Solve the problems now the solution for all these problems, so how can be used Hadoop as a solution for all the problems. We were discussing Hadoop it so firstly how to process Framework that allows us to store and process large data sets in parallel and distributed fashion. so to explain it further it is an open source Framework which is used for parallel and distributed computing where you can store and large data sets at the same time that is and Francis Hadoop status at the same time that is the charge that is the beauty of hello so it is not one particular software it is a framework. It is safe framework and it allows us to store and process large data sets ok and it helps you to do it in parallel as well which was the third one the third challenge that we saw that the data cannot be process faster because we keep you have a one master and mini slave architecture then it would be serving one by one only but here if you see here it is doing everything doing everything in parallel and in a distributed manner so that is distributed computing you know that the data is distributed across the different servers and then the processing happens so something like that so how does Hadoop solve a problem the main two components is HDFS map reduce which are part of which helps is too kind of small and big data from HDFS in Hadoop distributed file system which is Hadoop storage system the name implies it is a file system and it is not a database it is not a database on right so it is the file system like your Windows file system or your Linux file system it is also a file system but it is Hadoop distributed file system it the data is distributed data is not store in one Li one place it is distributed alright so because cause of distributed nature it allows to down any kind of data across the cluster the group of service bright so you can download any kind of data across the cluster using the HDFS connected servers can actually share the story on rent and this is to store the data and how do I process the data in a distributed manner is done using processing is done using map reduce so produce is used for parallel processing of data that is stored in HDFS so I have my data in HTML if I have to do some processing on HDFS I have to use map reduce to do it now let me explain you further on how this helps you to solve the problem write so two things to keep in mind now Hadoop is a frame which which allows to store and process La status in parallel and distributed and restoring is using HDFC and processing is using map .

V. WHAT IS MACHINE LEARNING?

As you know, we are living in a world of humans and machines, humans have been evolving and learning from past experiences for millions of years, on the other hand, the age of machines and robots has just started in today's vacuum. nowadays, the theses of the machine or robots are like them. it needs to be programmed before actually following its instructions, but what is the machine that started learning itself and this is where machine learning comes into the picture? Machine learning is at the core of many futuristic technological advancements in our world today, you can see various examples or implementations of Learning Machines all around us such as Tesla's Apple Siri robot Sophia Al and many more, so what is it exactly machine learning? Machine learning is a subfield of artificial intelligence that focuses on the design of systems from which you can learn and make decisions and predictions based on experience, which is the data in the case of machines, learning allows the computers act and generate data. Decisions driven instead of being explicitly programmed to accomplish a certain task, these programs are designed to learn and improve over time when exposed to new data, let's go ahead and discuss one of the biggest confusions of people in the world. of them, machine learning and deep learning are the same, you know what is wrong, let me clarify things for you, artificial intelligence is a broader concept that machines can perform tasks in a more intelligent way, it covers everything that it allows computer to behave like humans think of a famous Turing test to determine if a computer is capable of thinking like a human or not if you are talking to Siri on your phone and you get an answer, you are already very close to that, so This was about artificial intelligence now coming to the machine learning part, so as I already said that machine learning is a subset or current application of AI, it is based on the idea that we must be able to give the machine accessing the data and letting it learn from itself is a subset of artificial intelligence that deals with the extraction of patterns from the data set, this means that the most who can not only find the rules for optimal behavior

But they can also adapt to changes in the world, many of the algorithms involved have been known for decades, even thanks to advanced computing and parallel computing, now they can scale to massive volumes of data, so this was the part of machine learning. Now deep learning is a subset of machine learning in which similar machine learning algorithms are used to train deep neural networks to achieve better precision in cases where the former was not performing correctly in the process. Map. I hope now you have understood. that Al machine learning and deep learning are all different.

How do you solve a problem using machine learning?

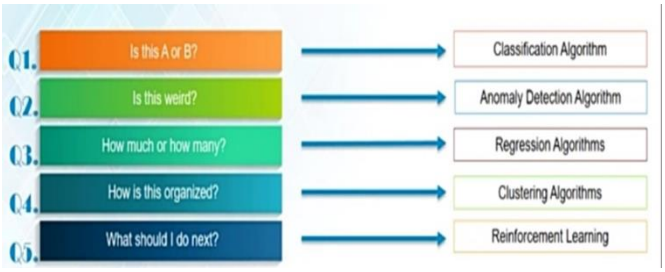


Figure 4: How solve a pro using machine learning

So whenever there is a problem, that boundary problem can be classified in five ways, so these five ways are like this: a or b is this, see how much or how many, how is it organized, and what should I do next to make its problem you may want an answer, but what could be is this a or b, since in your opinion you have a problem that is to ask, let's say you are differentiating between fruits in this an apple or a banana, so when you have this type of problem use classification algorithms the next category is this view analyze patterns, so when you have problems where you have to analyze the frame and where and you have to find a LOC anomaly or a strange one,

you are actually looking for an anomaly detection algorithm. The next category is how much or how many so you have to deal with the numbers. So when you want some numerical values, you want to get a certain value or for example what should be the minimum number of hours you have to spend to get a promotion. So when you have these problems and apply your expiration algorithm on that, then we have this organization. So when you have these types of questions, you use clustering algorithms because you are basically trying to figure out what the structure behind a certain data set is. So when you are trying to figure out this structure behind certain problems, you use clustering algorithms and then you have a category that says what should I do next, so well a decision has to be made and then algorithms are used for reinforcement learning.

Machine learning types

3 kinds of way so the 1st way Supervised Learning the 2nd way is Reinforcement Learning and then you have the Unsupervised Learning.

1. Supervised Learning:

The 1st kind of learning is called supervised learning. What is supervised learning so supervised if you concentrate on the word supervised, supervised means when monitoring someone or when you're constantly monitoring someone or making them understand some thing so you compare the scene with a classroom scene. So you sit in a class and the teacher explains you different concepts. The concept be learnt on your own that's why teachers is there so the way we teach machines in supervised learning is like this so we provide them with a particular set of inputs and we give the corresponding answer as for example if I am if I say that what are the parameters for deciding whether it will rain today or not so the humidity should be above some certain level temperatures should be above some certain level the brain should be in a certain direction and then if these scenarios are there it will rain right so we give a lot of inputs to the computer with this data and with each data we assign one it rains and zero it does not rain so if the temperature is high and the humidity is high and the wind is in a particular direction it will rain so we say 1 and if the humidity is low and the temperature is low as well so we say it will not rain so I will give it a certain input that today it's certified degrees Celsius the huge 97% the wind is in this direction will it rain so the machine will actually see ok so much up from my past experiences I saw that ok this was the temperature and the humidity was this it rained so it will compare that according to it and it will come up with the probability and hence come up with the solution or an answer whether it will rain or not so this is what supervised learning is so basically we are providing it with the Answers and the inputs as well all.

2. Unsupervised Learning:

Next topic is now unsupervised learning so what is unsupervised learning so you can understand on supervised learning by comparing it straightaway with supervised learning so in supervised learning like I said you were giving them ass well to input but an unsupervised learning you don't do that you just give them inputs now you are not giving telling your computer what will be the answer so what computer does is or what is it think logically the only thing computer can do with the inputs can do with the inputs is find a pattern behind it or find a structural in it right so this is what the computer actually does so in unsupervised learning what it does you provide inputs so for example I want my computer to I give my computer some inputs on fruit so I don't tell the computer what the flu pills actually but I give other parameters such as how big it is or what color it has say what is the taste of that fruit so when I give all these conditions or all these parameters to my computer so it groups the fruits basis on that so basically it will group on the basis of size it's a group

it be on the basis of taste it will group it be on the basis of color and then it shows us that data and then we can actually label ok so the size is big and color is this you this this fruit will be known as an apple now what kind of problems are actually there in unsupervised learning could be when we don't know whether there is correlation in the letter or whether there is a structure in the data for example so big data is nothing but a huge chunk of data so we don't know that ring that is Adam it's not structure so whenever we want to find a structure in data we use Unsupervised learning now it's the job of the algorithms to figure out what is if there's a pattern in the data and if at all there is a pattern it gives us that fact and hence decide how we can move ahead alright so if we have structure or if we have if you know how that data can be differentiated how it can be structured then supervised learning can be applied on it but If we don't know what that data structure is we use Unsupervised learning.

3. Reinforcement Learning

Reinforcement learning basically is when your computer is trying to take decisions. So what kind of problems can be included in reinforcement learning is say when you you're when you wanted to teach you computer how to play chess now you cannot tell your computer what to do because there are a lot of things that are like a zillion possibilities or zillion moves that can be done in chess. So you cannot tell each and every move to your computer but what you can tell is whether he did or wrong and that is what reinforcement learning. Is for another example that you can take is when you are training your dog all right so you cannot tell your dog what to do because, even not understand but you can actually reward your dog if it does and you can punish them who does wrong. So that same thing is actually applied in reinforcement learning as well so it basically the computers aim is to maximize rewards ,when it does the actions so it will come up with the solution which has the maximum rewards in place so we define if it does out sudden action you get a reward and then from its past experiences it understand. When I did this I got a reward so let me do something similar I will get mire rewards and that is what reinforcement learning.

VI. WHAT IS ALGORITHMS?

Very generic approach so if you want to interact with the computer and tell it how it should execute a particular task the way you do it is using a program right so what is the program a program is basically logic which is wrapped around a syntax a particular syntax which is specific programming language now this programming language could be anything it could be is to have a script it could be java it could be python it could be c it could be c++ whatever right but basic things doesn't seem that is the logic algorithm the logic remains the same in every language having said that what is this logic this logic is what an algorithm is alright so in simple words and algorithm is a step by step procedure towards solving a problem in the computer world all right so let's take an example to understand this thing which you exhaust discussed all right so let's take an example so this is a problem or this is an algorithm to print numbers from 1 to 20 so let's go step by step and understand what this algorithm is doing so this is the start passion we start over here and then we see that our algorithm is initializing a variable X to 0 So initialize a variable X to 0 and then we implemented it by 1 after that we are printing that variable and we are choking it whether it's less than 20 so if it's less than 20if it's true or if it's yes it goes back and increments the value gain by 1 otherwise if it's unknown it goes on and into program Since it's a yes right now it goes back and increments the value by 1 again so we are printed 1 and now we have increment the value of X by 1 again so now the value is 2 and then we print that value so we have now 1

and 2 on the board all then again it checks whether it's less than 20 if it's true it goes back again incremented by 1 now it's 3 prints see so this process goes on until the value of X reaches 20 so when the value of x is 20 it prints the value and it checks whether it's less than 20 which unknown and then it ends the program hence you have value which are printed from 1 to 20 so this is a step by step procedure for printing values between 1 to 20 and this is in similar way you would create other algorithms as well so as complex an algorithm can be it can always be represented using a flow chart.

Machine learning Algorithm:

1. Classification Algorithm

When you have a set number of outputs so basically question like this so is it cold outside today so it's answer will be either yes or no so you have only two outputs the outputs could be either yes but the output could be known all right or the next kind of question to be will you go to work today so its either a yes, no or maybe right so you will either go for work or you'll not go for work or you maybe you say maybe I'll go to work right but there is no other on service can come up right. So when you have these kind of conditions you come up with you solve it using classification algorithm .

Now, when you have two outputs like for example yes or no it is called two class classification but when you have more than two choices as in. In a second question we have yes, no or maybe so this is called multi class classification. So, whenever there is an output which is set as in it , is either true or false or 0, 1 or yes / no whatever. So if it's fixed you use classification algorithm this is basically the gist that you should get out of this any doubts and any doubt.

2. Anomaly Detection Algorithms

In these algorithm you analyze a certain pattern and you get alerted whenever there's an anomaly or something which is not usual which happens all. For example if as you can see in the fig that you have apart when you have some blue men and then 7 a red kind of person comes up so this algorithm will actually flag that person because, he breaks the pattern he is something which is not expected and he becomes an anomaly and this is what anomaly detection algorithms are about .Now what is the use case for anomaly detection algorithms it could be for example in credit card companies so in credit card companies each transaction of yours is monitored and whenever there is a transaction which is not usual . which doesn't match your daily transaction pattern you get alerted for it so they might confirm with you whether you only made this transaction all. So when you have these kind of problems you use anomaly detection algorithms to solve them.

3. regression algorithms

So like I said, whenever you have to get the value, use regression algorithms. So for example what the correct temperature will be for tomorrow, so whatever value comes out of this will be a number. Let's say I got to 23 degrees Celsius, I got that temperature using some formulas and I got that numerical value, so every time I come up with a numeric value of every time my problem demands that I have to get a mathematical value I go with algorithms of regression. So the second example of this could be whenever I want to give my customer a discount now, how much discount should I give that customer to get more customers and at the same time I don't go and lose? I'm still making a profit, so whenever this kind of problem comes up, I go ahead and use regression algorithms.

4. grouping algorithms

So clustering algorithms are basically used, which is why we discussed unsupervised learning, remember, so in unsupervised learning we have clustering algorithms. In which we try to establish a structure

so that we have some unstructured data that you want to make sense of, so what we do is use a clustering algorithm and if there is a pattern that we are calculating, we see that it comes up with that Pattern. and it shows us like this, for example, I send data to my computer and my data then apply a blocking algorithm on that, so this is the type of output that I will get, so it will categorize it and group A, group B and group C and then I can make a decision about what, what can I, what I want to do with this data that I have right, if this computer does not understand anything about this data, it does not understand, maybe it is of course. maybe it's food maybe it's money. So the computer doesn't understand, but what the computer understands are numbers and that's how it relates them to each other and creates groups. Now when you come up with groups, maybe you Wang will use someone else. I'll call it quantity and decide or come up with the solution, but this is what a clustering algorithm will give you.

5. reinforcement algorithms

Whenever you have a decision to make, and therefore whenever you have a decision to make and your decision is based on your machine's past experiences or whatever input you have given to your machine, you use reinforcement learning. Now, for example, whenever you are going to want to train your computer to play chess, it is usually application learning and when you have learned or when you have created a model for that and your game is actually being played by the computer. every decision the computer makes. it is also based on or also taken from reinforcement learning. The other example, like A to Z, was of a temperature control system where your system had to decide that it should increase the temperature or it should decrease the temperature. Therefore, whenever such problems exist, use reinforcement or reinforcement learning algorithms.

VII. CONCLUSIONS

Big data and machine learning algorithms are now being used in almost every industry of the economy. Central banks are also increasingly using big data for research purposes and to inform political decisions. In 2020, over 80% of central banks say they do used big data, up to from just 30% five years before. Among the institutions that currently uses big data, over 70% use them for economic research, while 40% say they use them inform political decisions. These numbers suggest that big data and machine learning offer many benefits questions and can help central banks fulfill their mandate. Current GDP and inflation or by examining spending patterns across regions and populations real-time subgroups provide just two of many examples. However, central banks also face challenges in unlocking the full potential of big data and automatic learning. A key topic of discussion is the availability of big data and tools for process it, store it and analyze it. The design of legal frame work or aspects of cyber security are also at the heart of central bankers' concerns. More practical problems they are budget constraints and difficulty in training existing staff or hiring new staff to work on big data issues. In fact, half of that there ported being interested in collaborating in many or more specific project, with there are three types of cooperation envisaged. First, by sharing knowledge among those institutions that have developed specific skills that they can be reused in other jurisdictions. These skills include general big data techniques (e.g. data visualization, network analysis, machine learning tools), more general information management problems (e.g. development of open-source code, data sharing protocols, encryption and anonymization techniques for the use of confidential information data) as well as specific applications more dedicated to the central bank community (e.g. suptech and regtech areas). Second, using big data to work on

global issues such as international spillovers, global and cross-border value chains

VIII. REFERENCES

[1]. Altig, D, S Baker, J Barrero, N Bloom, P Bunn, S Chen, S Davis, J Leather, B Meyer, E Mihaylov and P Mizen (2020): "Economic uncertainty before and during the Covid-19 pandemic", Journal of Public Economics, vol 191, September, pp 104–274.

[2]. Angwin, J, J Larson, S Mattu and L Kirchner (2016): "Machine bias", ProPublica, 23 May.

[3]. Amstad, M, L Gambacorta, C He and D Xia (2021): "Trade sentiment and the stock market: new evidence based on big data textual analysis of Chinese media", BIS Working Papers, no 917.

[4]. Andersen, A, E Hansen, N Johannesen and A Sheridan (2020): "Consumer responses to the Covid-19 crisis: evidence from bank account transaction data", CEPR Discussion Papers, no 14809.

[5]. Athey, S and G Imbens (2019): "Machine learning methods that economists should know about", Annual Review of Economics, vol 11, pp 685–725.

[6]. Bailey, M, R Cao, T Kuchler and J Stroebel (2018): "The economic effects of social networks: evidence from the housing market", Journal of Political Economy, vol 126, no 6, pp 2224–76.

[7]. Baker, S, N Bloom and S Davis (2016): "Measuring economic policy uncertainty", Quarterly Journal of Economics, vol 131, no 4, pp 1593–636.

[8]. Baker, S, R Farrokhnia, S Meyer, M Pagel and C Yannelis (2020): "How does household spending respond to an epidemic? Consumption during the 2020 Covid-19 pandemic", NBER Working Papers, no 26949.

[9]. Battaglini, M and E Patacchini (2019): "Social networks in policy making", Annual Review of Economics, vol 11, pp 473–94.

[10]. Boissay, F, T Ehlers, L Gambacorta and H S Shin (2020): "Big techs in finance: on the new nexus between data privacy and competition", in R Rau, R Wardrop and L Zingales (eds), The Handbook of Technological Finance, Palgrave Macmillan.

[11]. Bolón-Canedo, V, N Sánchez-Maroño and A Alonso-Betanzos (2015): "Recent advances and emerging challenges of feature selection in the context of big data", Knowledge-Based Systems, vol 86, September, pp 33–45.

[12]. Broeders, D and J Prenio (2018): "Innovative technology in financial supervision (suptech) – the experience of early users", FSI Insights, no 9.

[13]. Buch, C (2019): "Welcoming remarks International Seminar on Big Data", Building Pathways for Policy-Making with Big Data, Bali, 26 July 2018, also in IFC Bulletin (2015).

[14]. Cantú, C, S Doerr, G Cheng, J Frost and L Gambacorta (2020): "On health and privacy: technology to combat the pandemic", BIS Bulletin, no 17.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Phishing Attacks in Cyber Security

Prof. Pradip Sitaram Ingle[1], Ms. Pranita Shivkumar Murkar[2], Ms. Shital Prabhu Gadkari[2]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

In using various techniques of social engineering, the criminels font des ravages sur Internet et escroquent de nombreuses personnes de different manières. Cela met diverses communautes organizations, en danger. Therefore, it is important that people in those communities learn to protect themselves when active in cyberspace or when confronted with cyberspace-related technologies. In fact, training can play an important role in this regard and, consequently, help alter the unsafe behavior of many people. The goal of this research article is to determine whether simulating phishing attacks in conjunction with built-in training can help cultivate user resistance to "phishing attacks."

I. INTRODUCTION

In Phishing (pronounced as fishing), the social engineer tries to get the targeted individual to reveal personal information such as usernames, account numbers, and passwords. This is often done through the use of authentic-looking but fake emails from corporations, banks, and customer service personnel. Other forms of phishing try to get users to click on fake hyperlinks that will allow malicious code to be installed on the target computer without their knowledge. This malware will be used to delete data from the computer or use the computer to attack others Phishing is the act of attempting to acquire information such as username, password, and credit card details as a trusted entity in electronic communication. Phishing attacks have become more and more sophisticated and often mirror the target site transparently, allowing the attacker to observe everything while the victim is browsing the site and cross any limits above victim security. [1] As of 2020, phishing is by far the most common attack carried out by cybercriminals, with the FBI's Internet Crime Complaints Center recording more than twice as many phishing incidents as any other type of computer crime. [2] Attempts to prevent or mitigate the impact of phishing incidents include legislation, user training, public awareness, and technical security measures. [3]

Communications purporting to be from popular social web sites, auction sites, online payment process or IT administrators are commonly used to lure the unsuspecting public.

II. WHAT IS PHISHING?

On day to day base every one that used internet used a web application we can be used a web application through your apps on your smartphone or some application on your desktop or you mid be using a

web application. Now you can consider the keys where we using a web browser and you have to shop something online. Now you search for the product and you came cost in different websites and you like the products on to different website. So, one of this website is a very famous, very popular, very trusty ecommerce website and there is another website which is selling your product for the same price and may be for amore discount price, but it is not popular, you never heard that ecommerce website before. Now you chose to pay for a product online and for that you mid have to enter your credit card or debit card details. Now the question is which website would you trust more with your personal details with sensitive details that is a credit card details. Obviously you would deal less habitat well your entering the credit card details on the e-commerce website that is trustworthy and would be more hassidden would be worried while entering this details on to the website that your here in the name for the first site. This is because you don't trust the new website the other website the popular website you hear the name before you used before and you have trust factor with that website so psychology is what hacker take advantage of in phishing attacks, so hacker take advantage of this trust factor and they fake them cells as a trustworthy entity. To sale you are sensitive data and your personal data. So phishing is a attack of gathering sensitive information of a target such as username password email ID or other sensitive information maybe bank details your credit card, your debit card details etc, buy discussing as a trustworthy entity. So as a told you previously if you are entering any sensitive information maybe a card details on to a trustworthy website. You wouldn't be hesitate. So, if phishing attack hacker disguises him self as a trustworthy entity. Them his make you tricks you enter your sensitive information into that a fake web application. So, this is phishing.

2.1 How phishing works

Like you told you phishing is a web base applications and this is mainly used to cradiation . So we need a web application that is using a web server. Now every web application as connected to web server. When we have using a web application , what happened is there is some data and the packets, the information i.e. being send from your web application to the web server, and from the web server on the web application. Now, this is has an communication between web application and the web server happens. So, what happened the phishing attack is , the hacker disguises him self as this web server, so you think that your communication with the genuine with the actual web server but you reality you just communicating with the fake web server on the fake web application that the hacker has beld . And when you enter sensitive information on to this web server or this web application. The hacker used your prediction. This is how phishing work.

2.2 Step for phishing attack

1. Hacker creates a fake website.
2. F ake website is sent to the victim.
3. Victim enter credentials .
4. Hacker gets the credentials.

Step 1 : Hacker creates a fake website because, like I told you a phishing is a attack where a hacker disguises himself as a trustworthy entity. So, first he has to create a fake website, a fake of genuine website at to trick the victim to enter the credentials.

Step 2 : The next step is to send this fake website to the victim now, suppose a victim is trying to access a Facebook for example, if it goes yo the web application by him it's self may be he enter the URL of the websites on search

engine and then used the link go to that website. Than he does that he goes to the actual the genuine website and not to the fake website. So the second step is hacker has to send this fake website to the victim where, the victim enters the credentials.

Step 3 : The third step that happened is the victim he used that this fake website is a trustworthy website and enter credentials.

Step 4 : And finally the hacker gets the credentials.

2.3 Types of phishing

The attack can be carried out in several ways. drive for all types are similar. The only difference between Types are the number of targets and the mechanism used get data.

The different types of phishing attacks are [4]

- **Deceptive phishing :-**
  It reflective attached by we front emphases altimetry company. Once use of click the link off word credentials user personal information or login credentials are stolen. A fake email from a bank asking you to click a link and verify your account details is ones such a example.
- **Spear phishing :-**
  It target integer specific kind of profile. Attach a the social media and other websites. Forester customized attack email with target specific information that includes is name his position, his company to trick the recipient into believe in the mail her he gone into the confidential information. Take secret and even famashial gain.
- **Whaling :-**
  Her target is high post parson in an organization such as CFO, CEO. etc The high designating person more information, and Hence could be more damage into the organization this is the also known as CEO fraud.
- **Pharming :-**
  This is a form of online fraud where in bad actor relevant the clone or duplicate website to still information. Bad actor may used BNS redirection or occasionally MITM attack to execute it successfully.
- **Clone phishing :-**
  In this attack target are presented with a copy of ligament massage that they have regular earlier. But with a specific changes that the attacker are made in an attempt trap the target this change could be related to the malicious attachment invalid URL ext. Because communication was the part of the earlier communication this kind of attack are more successful or effective.
- **SMS Phishing :-**
  It is the combination of SMS and phishing. It is an act of commentating at text message fraud try to lure the victim into revolution account information or installing a software. The presenting to be limit emptily.
- **Voice phishing :-**
  That is voice phishing and vising. Her people are treat into giving money or veiling personal information vising frequently in was a criminal patent into the pre- presentative of bank or organization. It is a fondle on impact is, of existing sensitive information over the phone. Here, the hacker uses the VUIP technology so have to prove the caller ID, so that it's look like the call come from the authentic source.
- **Baiting :-**
  It realizes an quercetin or grid of the victim. Better offer users free download an using long in credential also and infected storage media like CD ROM or USB flash drive line unattended are used to live of the victim. Once the parson insert this drive into the computer there computer get infected with tandoors and backdoor emprise created in the victim computer.

- **Search Engine Based Phishing :-**
  Same phishing camsondboll search engine whether uses are directed to product are its. Its may offer low cost product or services. When the user try to by the product by entering the credit card details there collected by the phishing side. There are many fake bank websites of reacting credit card or loans to user at a low rate but actually they are phishing site

2.4 Phishing Technique

- Link Manipulation
- Filter Evasion
- Website Forgery
- Phone Phishing

2.5 Effect of Phishing

- Internet fraud
- Identity theft
- Financial loss to the original institutions
- Difficulties in law enforcement investigations
- Erosion of public trust in the internet

2.6 Defend against phishing attack

- Preventing a phishing attack before it begins
- Detecting a phishing attack
- Preventing the delivery of phishing messages.
- Preventing deception in phishing messages and sites.
- Counter measures

Interfering with the use of compromised information.

III. ANTI-PHISHING TECHNIQUES

In [5], the authors used text mining to extract distinct features of emails. Emails can be fraudulent or real emails To better detect the attack. The strategy used was Initial conversion of email to vector representation Followed by feature selection techniques for classification. The evaluation was performed using accumulated data sets From the Ham Corpus of Spam Assassin project (legal email) and publicly available Phishing Corpus E-mail).

The study carried out in [6] used a unique technique from Natural Language Processing (NLP) to determine if mail has been malicious or not. In this article, they extracted and compared common features using NLP tools. PhishNet-NLP used natural language techniques with all information present in an e-mail, namely the header, links and text in the body. PhishSnag used information from e-mail to detect phishing. Phish-Sem used NLP and statistical analysis on the body to label the mail as phishing or not phishing.

The authors talk about reusable components for anti-phishing component layer in [7]. These reusable components are used to convert web pages to feature vectors using heuristic methods and external repositories. The finished line the vectors which provide input to these vector machines cause the support vector machine. Thanks to the training provided by these inputs, the classified and determined support vector machine various web pages as legitimate or a phishing web page. This has been experimented with the mixture of heuristics to identify a phishing web page. A more advanced filtering and classification technique has been used in [8]. In this article, the authors tested URLs and checked whether it was malicious or not. They used a automated approach to detect phishing. He had two phases- Pre-filtering and classification phase. In pre-filtering phase, the URL was matched against a blacklist using the domain part of the URL. If the

URL was present in this list then it was classified as malicious and would not be move on to the classification phase. In the next phase, two the main characteristics were checked for consistency- the RU and the Position of thedomain. Based on the result of the classification phase, the URL was classified as malicious or non-malicious.

In [9], the anti-phishing technique was developed with the help of the advanced heuristic approach. In this technique, when suspicious site was encountered, it was immediately updated in the blacklist. If a legitimate website is found, it updates the same in the whitelist. Therefore, when the user open a website, it was first checked if the website was a phishing or non-phishing website and, accordingly, provided access to the same. This technique uses the PHP programming language with a database to maintain the two lists.

The method used in [10] is slightly different from the others technical. In [8], they classify emails as unwanted or non-unwanted based on spam filter. When an email arrives at mailbox, Spam filtering is performed on the basis of the reputation of the URL present in the to post. If the URL looks dangerous or suspicious, the filter marks the mail as "junk". URLs in mail are disabled and the mail is then moved to junk mail. If the mail is genuine, the mail is moved to the inbox for the user to open without issue.

The method of extraction and classification was further developed in [11]. In this article, the vulnerabilities have been differentiated into three categories according to the structure of the email. The three categories were the content of the page vulnerability, domain vulnerability and code script vulnerability. The evaluation model used was Anti-Phishing Effectiveness evaluation model (APEE model) that is used analyze the effectiveness of anti-phishing mechanisms that have been implemented. The reputation of vulnerabilities in all three categories are tested, which helps to determine whether the email is a phishing email or not.

Example:

1. Petya Ransomware Attack Date : 27-28 June 2017

Victims: Ukraine (80%) ,Germany (9%), Russia ,Italy, France, UK ,Poland ,US

Attack Type: Cyber Attack - Ransom ware

Attacker: Russia (according to Ukrainian authorities and the CIA)

Intention: Politically motivated attack against Ukraine

Results: More than 80 companies (Several Ukrainian ministries, banks and metro systems) including national bank of Ukraine

Description:

On 27 june 2017 a major cyber attack occurs on Ukraine and it was a ransomware name "petya". Same cyber attack reported from Germany , Russia US .

The main target of this cyberattack is Ukraine. More than 80% companies in Ukraine were affected with petya.

80% infections were reported on Ukraine and 9% infections were reported on Germany.

There was enough evidence that Russia is responsible for this and it was done out of political motivation against Ukraine.

This ransomware target on companies power grid s, bus stations, gas stations, airports, and banks.

This attack were originated from an update of Ukrainian tax accounting package called MeDoc (M.E.Doc) developed by Intellect Service.

MeDoc is common among the Ukrainian accountants.

The impact occurs on around 400,000 customers and 1 million computers across Ukraine.

What is petya & How does it work

Petya is an encrypting ransomware that is first discovered in 2016.

This targets Microsoft Windows based systems infected on the master boot record (MBR).

Then execute a payload that encrypt the file systems of hard drive and prevent Windows booting.
To regain the access to file system user needs to pay a ransom with Bit coin to the attacker.
Used for performing a cyder attack on Ukraine

There are four steps in petya.

1. Prepare - Attack begins with compromising the MeDoc application when an organisation updated the application the Petya code will be inserted.
2. Enter - When customers update MeDoc software Petya code runs on an enterprise host and spread over the enterprise network.
3. Traverse - There are two path malware traverse.
   *Exploitation - Exploiting the vulnerability SMBv1 (MS17_010).
   *Credential theft - Impersonate any account that currently logged in to the account.
   *Petya only compromised accounts witch has an active session (credentials loaded into LSASS memory)
4. Execute - Petya then reboot and start the encryption process. Then user can display screen text that is a ransomware.

Impact:

During the attack, radiation monitoring system at Ukraine's Chernobyl Neclear Power Plant went offline.
Inflected to Serveral Ukrainian ministries, banks,metro systems and state-owned enterprises (Boryspil International Airport, Ukrtelecam, Ukraine's electricity company's computers also went offline (But Company continued to fully operate without using computer). Inflected computer data were overwritten and permanently damaged.
The total estimated damage was more than $10 billion

Mitigation:

Patch Management: lmplement an emergency patch program and ensure that all Windows systems are receiving security patches from Microsoft and other vendors on a frequent basis. This patch relevant to fixing the Eternalblue vulnerability is MS17-010.
Host Based Firewalls: Consider applying firewall rules at the host level (Windows firewall) which prevent unnecessary system to system communication (making it more difficult for Worms to propagate).
Network Segmentation: Properly segment networks and apply routing and firewall rules which create security zones within your network, limiting the attack surface of malware was introduced. Use Supported Operating System: Ensure all operating systems currently being ran by the vendor (Windows XP and Server 2003 are no longer supported or receiving security updates).
Properly Manage Backups: Verify that backups are not stored within network attached directories that might be susceptible to being infected by a Worm (End up being encrypted as part of the ransomware attack).

Cognizant

One of the worst largest provide sophisticate services sad they become ogecting of a ransomware attack. Then thus maze ransomware attack. That has goes description of yields client. Company realized a official statement on its website and its states. Cognizant conform there security incident involving our internal system and cozing services disruption for some of our clients. It the result of a maze ransomware attack. They also said cognizant conform that it is taking spoke of the incident and already started communicating with all there client on this

maze ransomware attack. The reported maze attack on cognizant is horizon as it is not like a typical ransomware other than encrypting data it is also to spread accrosed network infecting and increasing every computer on its swag and it can also oxidation data of the attackers.

“The Maze attack reported on Cognizant is worrying, as it does not look like typical ransomware. Besides data encryption, it is capable of propagating over a network, infecting and encrypting every computer in its path, and it can also exfiltrate data to attackers,”said Saket Modi, CEO of Lucideus, a corporate cybersecurity platform company.” [12]

“Although Maze operators denied the attack, it was always classified as Maze because indicators of compromise listed included the IP addresses of servers and file hashes, which are known to be used in previous attacks by the. Maze actors.”[12]

At a time when more than 90% of IT service company employees around the world work from home, such an attack indicates a worrying trend. While there has been a significant increase in phishing attacks in the form of Covid-19-themed websites, the ransomware attack appears to be the most serious form of such attempts at this point. [12]

IV. CONCLUSION

Phishing is a technique to collect confidential information about the target via malicious links and emails. It is one of the most dangerous cyberattacks that occur in organizations, personal devices, etc. It is often difficult to distinguish between genuine emails and phishing emails. exist various methods that can be used to avoid this attack. Regular updating of anti-phishing tools and platforms can prove to be very powerful. This study provides a perspective of phishing, the mechanism of the attack, various ways it can occur and possible solutions to overcome them.

V. REFERENCES

[1]. Ramzan, Zulfikar (2010). "Phishing attacks and countermeasures". In Stamp, Mark; Stavroulakis, Peter (eds.). Handbook of Information and Communication Security. Springer. ISBN 978-3-642-04117-4.

[2]. "Internet Crime Report 2020" (PDF). FBI Internet Crime Complaint Centre. U.S. Federal Bureau of Investigation. Retrieved 21 March 2021. "Internet Crime Report 2020" (PDF). FBI Internet Crime Complaint Centre. U.S. Federal Bureau of Investigation. Retrieved 21 March 2021.

[3]. Jøsang, Audun; et al. (2007). "Security Usability Principles for Vulnerability Analysis and Risk Assessment". Proceedings of the Annual Computer Security Applications Conference 2007 (ACSAC'07). Archived from the original on 2021-03-21. Retrieved 2020-11-11.

[4]. Gaurav, Madhuresh Mishra, Anurag Jain - Anti-Phishing Techniques: A Review, International Journal of Engineering Research and Applications (IJERA), vol. 2, pp. 350- 355, April – 2012.

[5]. Masoumeh Zareapoor, K.R. Seeja, Text Mining for Phishing E-mail Detection, Intelligent Computing, Communication and Devices: Advances in Intelli-gent Systems and Computing, vol. 308, pp. 65-71, August 2016.

[6]. Rakesh Verma, Narasimha Karpoor, Nabil Hossain and Nirmala Rai - Automatic Phishing Email De-tection based on Natural Language Processing Techniques, Research Gate, 2016.

[7]. Anna L. Buczak, Erhan Guven - A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection, IEEE Communica-tons Surveys & Tutorials, vol. 18, pp. 1153-1176, June – 2016.

[8]. Yi-Shin Chen, Huei-Sin Liu, Yi-Hsuan Yu and Pang-Chieh Wang, Detect Phishing by Checking Content Consistency, IEEE, 2017.

[9]. Okunoye, O.B, Azeez, N.A, Ilurimi F.A - A Web Enabled Anti-Phishing Solution Using Enhanced Heuristic Based Technique, FUTA Journal of Re-search in Sciences, vol. 13 (2), pp. 304-321, Octo-ber – 2017.

[10]. Xavier Joseph, Mitchell Aime M, Tsang Brian J, Herbert, George A, Savastano, Hernan I, Khandel-wal Lubdha, Pengelly Robert C. J, Novitskey Robert, Grant Stanley - Anti-phishing protection, United States Patent 10,069,865 B2, Sept 4th, 2018 Sankhwar S., Pandey D., Khan R.A - A Novel Anti-phishing Effectiveness Evaluator Model, Smart Innovation, Systems and Technologies, Springer, vol 84, Cham, 2018.

[11]. "PM Narendra Modi pushes for use of technology in the 'era of Covid-19'"

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Malaria Detection Using Supervised Learning

Prof. S. T. Sawle[1], Ms. Samruddhi Sudhir Kavimandan[2], Ms. Prachi Pradip Narkhede[2]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

Malaria is a deadly, infectious and life-threatening mosquito-borne blood disease caused by Plasmodium parasites. The conventional and most standard way of diagnosing malaria is by visually examining blood smears via microscope for parasite-infected red blood cells under the microscope by qualified technicians. This method is inefficient and time consuming and the diagnosis depends on the experience and the knowledge of the person doing the examination. Automated image recognition technology based on image processing has previously been applied to malaria blood smears for diagnosis. However, practical performance has so far not been limited. It gives us all the impetus to make the diagnosis and diagnosis of malaria faster, easier and more efficient. Our main goal is to create a model that can detect cells from multiple cell images of a thin blood smear on a standard microscope slide and classify them as infected or not by early or effective testing using image processing. And also classify infected cell images using machine learning. Key Words: Malaria, Falciparum, Watershed, Morphological Segmentation, Edge Detection, and Segmentation.

I. INTRODUCTION

Malaria is a deadly, infectious disease caused by the Plasmodium parasite which is transmitted by the bites of female Anopheles mosquitoes. According to the World Malaria Report 201 Report published by WHO [1], an estimated 10,000,000 malaria-related deaths were recorded last year. The disease is curable but early diagnosis is key. Existing methods used to detect malaria include microscopic examination of infected cells in the laboratory. This method is both expensive and tedious. The WHO African region recorded approximately 100 percent of all malaria cases in 201 in. The region has one of the highest per capita incomes in the world. This model offers a fast, low-cost and reliable alternative to micro-testing for malaria.

1.1 Problem Statement :

We propose an image processing model for detection of malaria infected cells. We use image processing techniques to detect parasite-infected red blood cells in thin smears on standard microscope slides. The most widely used present day method is analyzing thin blood smears under a microscope, and visually searching for contaminated cells. A clinician manually counts the number of parasitic red blood cells - sometimes up to 5,000 cells (according to WHO protocol) [2].

Malaria could be forestalled, controlled, and relieved all the more adequately if an increasingly precise and

effective symptomatic technique were accessible. We have used image processing to identify near malaria contaminated cells. And to classify the state of malaria, whether it is falciparum or not, it is the most deadly stage in malaria or non-falciparum, for which we use machine learning technology.

1.2 Scope of the Project:

Malaria screening from thin film blood smear images demands the separation of single blood cells from microscopic blood slide images that can be taken from a pathologist and the dataset does not have cell images that do not have division. Therefore, in the proposed method the partitioning is done using different image processing techniques. Edge detection techniques and segmentation techniques used in this system overcomes the issue of overlapping of cells by eliminating the noise and finding the discontinuities of the cells. It differentiates between cells and identifies infection in cells using morphological segmentation. Also, all the images are raw and their intensity is different and it is very difficult to detect cells and infection as there is no uniformity in all the images. To overcome this problem, the proposed method uses histogram matching where all images are standardized and have the same intensity which increases the level of accuracy.

Frankly, there are 3 types of machine learning algorithms

1. Supervised Learning

How it works: This algorithm contains a target / result variable (or dependent variable) that must be inferred from a set of predictive variables. Using a set of these variables, we generate a function that maps the desired output. The training process continues until the desired accuracy level on the model training data is achieved. Examples of supervised education: regression, decision tree, random forest, KNN, logistics regression etc...

2. Unsupervised Learning

How it works: In this algorithm, we have no target or result variable to predict / predict. It is used for population clustering between different groups, which are widely used to divide customers into different groups for specific interventions. Examples of ineffective learning: ri priori algorithm, k-means.

II. LITERATURE SURVEY

The Plasmodium parasite is parasitic malaria that invades red blood cells (RBCs) and is transmitted by mosquitoes. Neurol networks are used to analyze the potential of RBC and parasitic ethin blood smears. In, the loaded KNN (K-close neighbour) algorithm is trained by the options taught in the Abuse Theorem Picture Component Classifier, which aims to spot pixels. To identify multi- category parasites according to the lifecycle stage and its species

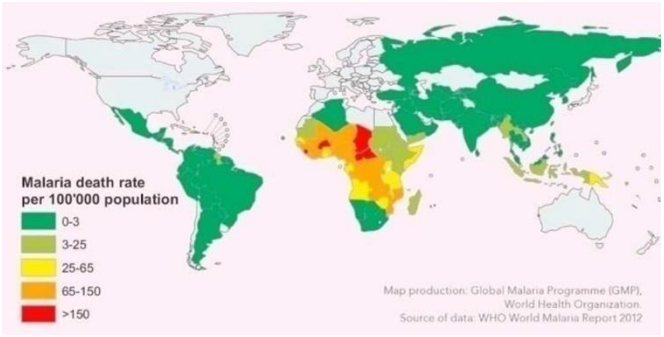


Fig1 Worldwide malaria death rates

Basic thresholding is accomplished using a bar graph-based procedure to detect the presence of plasmodium in blood smears. It is very important to prepare the smear, as these changes can lead to changes in imaging conditions. The misbehavior of the overlapping RBC was the morphological operator. Deformed square measurement reported by analyzing cells containing cells where the image of authenticity is binaryized and using a vague live technique. Further, these tagged cells have some properties such as color, size, and choice as a rank neural network mistreatment using square-sized platelets, leukocytes,

and corpuscles classified designs. Over the past few years there has been a good deal of agreement to develop new methods for detecting protozoal infections, including fast substances, fluorescent research detection methodology, and PCR (polymerase).

Chain reaction) The process of finding specific sequences of macromolecules. Nonetheless, the method of identifying light weight research is the most common and commonly used technique.

The research will differentiate between species categories, quantify parasitism, and monitor the different acute stages of the parasite. But this method requires trained technicians and it is a time consuming process and the ultimate precision of identification also depends on the skills and abilities of the scientists and the amount of time they spend learning each slide. Malaria is a life- threatening disease and scientists around the world are interested in analyzing it. Previously, a large number of protozoal infections were diagnosed in a laboratory setting that required good human experience. Automated systems such as hope for machine learning techniques have previously been studied to overcome this negativity.

During the study of this domain, technology gave a lot of thought to hand-crafted options when reporting. For example, SVM and Principal Part Anal Analysis (PCA) were applied for morphological factors and classification purposes for characterization. However, the accuracy achieved by these types of models is less than that of the in-depth learning-based techniques studied recently.

What is special is that during this work, we have a tendency to propose in-depth models that bring home the Bacon classification performance, such as the highly accurate deep learning already reported. In addition, our models measure economically in terms of required processing resources and fail to perform efficiently on a good mobile device, as well as with the square measures offered at a low cost. [3]

III. MACHINE LEARNING TECHNIQUES

3.1. Supervised Learning

Supervised education is the most common sub-branch of machine learning today. In short, new machine learning practitioners will begin their journey with supervised learning algorithms. So, the first supervision in this three post series will be about education.

Supervised machine learning algorithms are designed for example learning. The name "supervised" education comes from the idea of training this type of algorithm as teachers supervise the whole process.

When training supervised learning algorithms, the training data will include inputs connected with the appropriate output. During the training, the algorithm will search for patterns in the data related to the desired output. After training a supervisory learning algorithm will take new unseen input and determine which labels will be classified based on previous training data. The purpose of the supervised learning model is to predict the appropriate label for the newly introduced input data. In its most basic form, supervisory learning algorithms can simply be written as:

$Y=f(x)$

Where y is the predicted output that is determined by the mapping function that assigns the square to the input value x. The task used to connect the input features to a prediction is created during training by a machine learning model.Supervised learning can be split into two subcategories:

Classification and regression

3.1.1. CLASSIFICATION

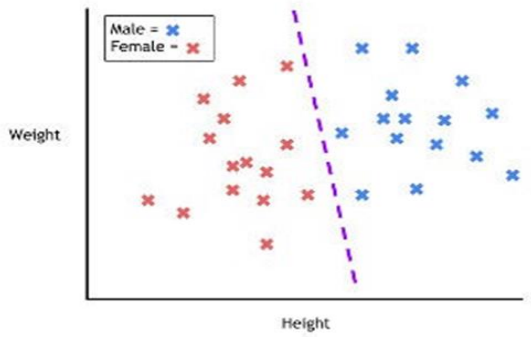


Fig2

During training, a classification algorithm will be given data points with an assigned category. The function of the classification algorithm is to then take the input value and assign it a class or category based on the training statistics provided to it. The most common example of classification is determining whether or not email is spam. With two classes to choose from (spam or no spam), this problem is called binary classification problem. The algorithm will be provided with training data containing non-spam and non-spam emails. The model will find features in the data that are either consistent with the range and create the mapping function mentioned earlier: Y = F (X). Then, when an unseen email is provided, the model will work to determine if the function is spam. Classification problems can be solved with a number of algorithms. Which algorithm you choose to use depends on the data and the situation. Here are some popular classification algorithms:

- Support Vector Machines
- Decision Trees
- K-Nearest Neighbor
- Random Forest

3.1.2. REGRESSION

Regression is a predictive statistical process where the model attempts to find the important relationship between dependent and independent variables. The goal of the regression algorithm is to predict continuous numbers such as sales, revenue, and test scores. The basic linear regression equation can be written as: Y=w[0]*x[0]+w[1]*x[1]….+w[i]*x[i]

Where x[i] is the feature(s) for the data and where w[i] and b are parameters which are developed during training. For a simple linear regression model with only one feature in the data, the formula looks like this:

$$\hat{y} = wx + b$$

Where W is the slope, x is the only feature and B is the y- intercept. Familiar? For simple regression problems like this, the model estimates are indicated by the best fit line. This aircraft will be used for models that use two features. Finally, hyperplanes will be used for models that use more than two features. Imagine being able to determine a student's test grade based on how many hours we studied during the exam week Let the plotted data with the best fit line look like this:

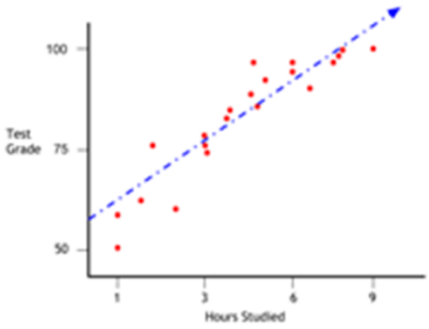


Fig3

There is a clear positive relationship between the hours studied (independent variables) and students' final test scores (dependent variables). Given the new input, the best fit line can be drawn through the data points to show the forecast of the models. Say we wanted to know how well a student would do with five hours of study. We can use the best fit line to estimate test scores based on other students' performance.

There are many types of regression algorithms. The three most common are listed below:

- Linear Regression
- Logistic Regression
- Polynomial Regression

3.2. Unsupervised Learning

Unsupervised learning is where you only have input data (X) and no corresponding output variables.
The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.
These are called unsupervised learning because unlike supervised learning above there are no correct answers and there is no teacher. Algorithms are left to their own devices to find and present interesting structures in the data.
Unsupervised learning problems can be further grouped into clustering and association problems.

3.2.1. Clustering:

A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior. Clustering is a data mining technique that groups unlabelled data based on their similarities or differences. Clustering algorithms are used to group information on raw, unclassified data objects by structures or patterns. Clustering algorithms can be classified into several types, especially unique, overlapping, hierarchical, and potential.

3.2.2. Association:

The problem with learning association rules is where you want to find rules that describe a large portion of your data, just as people who buy X also buy y. An association rule is a rule-based method for finding the relationship between variables in a given dataset. . These methods are frequently used for market basket analysis so that companies can better understand the relationship between different products. Understanding customer usage habits enables businesses to develop better cross-selling strategies and recommendation engines. Examples of this can be found in Amazon's "Customers Who Purchased These Items" or in Spotify's "Discover Weekly" playlist. While some different algorithms are used to generate association rules such as apriori, eklat, and fp-growth, the ri priori algorithm is the most widely used.

IV. RANDOM FOREST IN MACHINE LEARNING

Random forest algorithm is most famous and easy to use machine learning algorithm based on ensemble learning. In this article you will learn how this algorithm works,how it’s efffecient comparing to the other algorithms.

4.1. What is Random Forest in Machine Learning?

Random Forest is a supervised machine learning algorithm that can be used to solve classification and repression problems. However, most of them prefer classification. It is named as a random forest because it combines multiple decision trees to create a “forest” and feed random features to them from the provided dataset. Instead of depending on an individual decision tree, the random forest takes prediction from all the trees and selects the best outcome through the voting process.
Now, the question arises why do we prefer random forests over decision trees. So, individual plants are more useful but random forests can reduce this problem by averaging the estimated results on each plant.

The basic idea of random forestation should be explained here. Let's dive deeper into this and understand how this algorithm works.

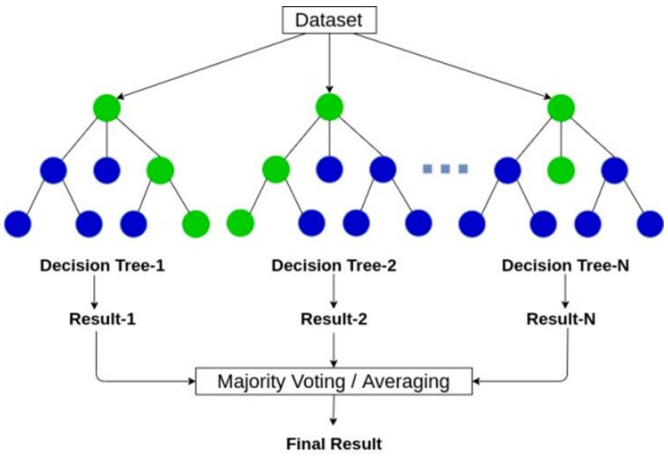


fig. 4 Working in Random forest

4.2. 2How it works?

We can understand the working of a random forest with the help of the given steps:-

- Select random patterns from the dataset provided by Data.
- Selected Create a decision tree for each selected pattern. You will then get approximate values for each tree created.
- Then for each predicted result voting will be done.
- In the end, the algorithm will choose the result (predicted) with majority votes.

V. ADVANTAGES AND DISADVANTAGES

5.1. Supervised Learning

Advantages

1. An example of linear regression is easy to understand and fairly straightforward. This can be generalized to avoid overzealousness. Furthermore, linear models with new data can be easily updated using a statistical gradient
2. The use of well-known and labelled input data makes supervised learning produce a far more accurate and reliable than unsupervised learning. With the access to labels, it can use to improve its performance on some task.
3. Efficient in finding solutions to several linear and non- linear problems such as classification, robotics, prediction and factory control. Neurons able to solve complex problems by hiding leather (Satya and Abraham, 2013).

Disadvantages:

1. Takes a long time for the algorithm to compute by training because supervised learning can grow in complexity. Therefore, since most of the data in the world is unlabeled, results are not actually obtained, performance is very limited.
2. Performs poorly when there are non-linear relationships. One of supervised learning method like linear regression not flexible to apprehend more complex structure. It takes a lot of computation time and also difficult to append the right polynomials or interaction terms.
3. Its not cost efficient if the data keeps growing that adds to the uncertainty of data labelling to predefine outputs. Example, It is costly to manually label an image dataset, and the most high quality image dataset has only one thousand labels, according to (Ankur A., 2018).

5.2. Unsupervised Algorithm

Advantages:

1. Lets algorithm to refer back for patterns that has not been accounted previously, therefore resulting the freedom of learning direction for the algorithm in unsupervised learning (Kohonen and Simula, 1996).

2. Excels at problem where insufficient labelled dataset or identifying unknown pattern or constantly evolving. learning the concealed pattern of the data it has trained on. Makes previously unmanageable problem more solvable and more agile at finding hidden structure in past data and future prediction (Ankur A., 2018).
3. Simplified human task of labelling by grouping similar object and differentiating the rest. This group of datasets will then be labeled one after the other instead of being labeled (Ankur A., 2018).

Disadvantages:

1. According to (Stuart and Peter, 1996) a completely unsupervised learner is unable to learn what action to take in some situation since it not provided with the information. The goal of unsecured learning, which is not accessible to any output, is simply to find a pattern in an available data feed.
2. Quite slow and consumes large resource memory, therefore harder to scale to larger datasets. Furthermore, it assumes that only the original clusters in the dataset are globe-shaped.
3. The outcomes are not that accurate due to it is mostly about prediction. In addition, we do not know the number of classes, therefore the results are not certain.

Unsupervised learning is less adept to solve narrowly defined problem (Silvia, 2018).

VI. RESULT

Here Random Forest algorithm used in our topic becaused it has more acuuracy than others.
See the table below

Algorithm	Accuracy	Precision	Recall	F-Score
Decision Tree	0.965	0.526	0.529	0.527
KNN	0.940	0.465	0.278	0.348
Linear Regression	0.943	0.375	0.000	0.000
Naive Bayes	0.858	0.271	0.873	0.414
Random Forest	0.965	0.775	0.553	0.645
Extra Tress	0.956	0.837	0.298	0.440

Performance of different Machine learning algorithm

Python has various in-built libraries used specifically for implementing machine learning algorithms. Fortunately, we don't have to write code to implement complex parts, we can accomplish things just by importing these libraries.We used Random Forest classfier algorithm here becaused it has more accuracy than other algorithms.

VII. CONCLUSION

We have proposed a Malaria parasite detection method using a shallow machine learning algorithms. This method of detecting the malaria parasite can be very useful to health workers in countries, where there is less number of trained laboratory experts and lack of resources. In the present work, we divided the image into patches and analysed based on the presence or absence of malarial parasite. To accomplish this, we have used various classical machine learning algorithms such as AdaBoost, Decision Tree, KNN, Random Forest, etc. The accuracy of our model assists the laboratory technicians in decision and classification framework

may be sufficiently general for other diagnostic tests like hem parasites, worm infestations, or tuberculosis

VIII. REFERENCES

[1]. World Health Organization, Malaria, https://www.who int/newsroom/factsheets/detail/malaria-report-2019

[2]. World Health Organization, Malaria, https://www.who int/newsroom/factsheets/detail/malaria(2018)

[3]. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Aspect Based Sentimental Analysis for Reviews on E-Commerce Platform for Business Improvisation

Shraddha H. Ingle[1], Prof. Mayur S. Burange[2], Dr. Ajay B. Gadicha[3]

[1]M. E. Scholar, Department of Computer Science and Engineering, P. R. Patil College of Engineering and Technology, Amravati, Maharashtra, India

[2]Assistant Professor, Department of Computer Science and Engineering, P. R. Pote (Patil) College of Engineering and Technology, Amravati, Maharashtra, India

[3]Head of Department, Department of Computer Science and Engineering, P.R. Pote (Patil) College of Engineering and Technology, Amravati, Maharashtra, India

ABSTRACT

As day-by-day large amount of data has been generating on digital platform, using this data provided by customers from all over the world which gives a positive hope for product manufacturer to make their product more consumer friendly and make consumer happy to buy it. When customer provide their opinion on any social media or e- commerce site their personal views on products may affect the way of looking towards product quality of the other customers. Sentiment analysis is a way to help product manufacturer to know about customer's opinion about the product. If some customer wants to know only about some feature of product, then aspect based sentimental analysis is the way to help those customers. Reviews is the one of the important type of data which is available in the form of reviews on various platform, here we take reviews from E-commerce platform to perform aspect-based sentiment analysis for products so that manufacturer will be able to take appropriate decision and provide customer suitable recommendation for products on the basis of reviews analytics.

Keywords: Sentimental Analysis, Aspect Based Sentiment Analysis, Naive Byes algorithm.

I. INTRODUCTION

Since last decade online purchasing has been growing rapidly and during this pandemic of covid 19 most of our life is dependent on online platform whether it is work or purchasing the products. As buying and selling has grown vast and generates lots of data. This data has been widely used by various businesses to develop and maintain customer- business relationship very well and to provide user friendly products. As now most of the things are going virtually customers opinions on various services and products has been shared online on various platform. Here we will be performing aspect based sentimental analysis system which will work for better performance by looking into reviews from various customers on e-commerce platform.

Sentiment analysis is the way to help to those producers who wants to develop or upgrade their products by knowing about their products advantages

and disadvantages from its users which is available in the form of reviews on various platform by using those comments as data and performing sentimental analysis on the data will be beneficial for those businesses and also for users to know about products from reading the previous reviews provided by previous customers.

Sentiment analysis, also opinion mining is the field of computational study that analyses people's opinions expressed in written language, where focus of research is on the processing of text in order to identify opinionated information [1]

II. RELATED WORK

(Duc-Hong Pham , Anh-Cuong Le, 2018) In this paper they propose a novel multi-layer architecture for representing customer reviews. They observe that the overall sentiment for a product is composed from sentiments of its aspects, and in turn each aspect has its sentiments expressed in related sentences which are also the compositions from their words.[3]

(Wenhao Zhang, Hua Xu , Wei Wan , 2012)In this work, they propose an expert system Weakness Finder by analysing the customers reviews on the influential web communities with aspect based sentiment analysis, it can help the cosmetic manufacturers to find the weakness of the products in order to improve their products. The Weakness Finder system can help to identify the features, and group the features into different aspects by using explicit and implicit features grouping methods, then judge the polarity of each sentence by using sentence-level sentiment analysis [5]. Sentimental analysis is defined as the analysis of opinions, thoughts, sentiments, and subjectivity of text are given [4].

(Fu Xianghua, Liu Guo, GuoYanyan, Wang Zhiqiang, ,2013) Propose an unsupervised approach to automatically discover the aspects discussed in Chinese social reviews and also the sentiments expressed in different aspects [6]

Sentiment Analysis (SA) is a continuous field of research in content mining field. SA is the computational treatment of feelings, estimations and subjectivity of content. This review paper handles a complete diagram of the last refresh in this field. Many as of late proposed calculations' improvements and different SA applications are examined and exhibited quickly in this review [2]

III. PROPOSED WORK

In the proposed model sentiment analysis will be perform for various reviews on e-commerce platform which will help business to know about user's opinions and be able to provide preferable recommended product for users. It will provide feature or aspect-based search for user so that they will know better about customers demand on various aspects. Till now various models focuses solely on business improvement perspective so that business can have growth but customers view is also equally important if manufacturer wants to achieve better improvement. So along with business platform, customer should also be happy to buy product and should not have to take hectic efforts of going through large no of reviews by reading them and to help those customers and manufacturers aspect-based sentiment analysis will help by reducing these efforts.

IV. RESULT ANALYSIS

Most of the models of sentiment analysis use SVM and Naïve Bayes algorithm for classification. In order to increase their accuracy and efficiency.[1] Here we use probabilistic approach by using naïve bayes algorithm by calculating probability of positive and negative reviews out of total reviews will give appropriate value whether the product is really getting its value in market or lacking behind

anywhere so that there should be chance to improve it and this all will be possible by using sentiment analysis on reviews by using naïve bayes algorithm. Here we take reviews from Kaggle dataset of reviews of women's clothing and dataset of reviews on mobile in which unstructured reviews will be process and after performing sentimental analysis it shows number of reviews according to probability of its positive reviews like good reviews, better reviews, bad reviews and negative reviews in bad reviews. It will also provide aspect-based search for customers so that if customer wants to know about product features, they will get it without any hectic search of previous reviews.

(a)→Actual Word Count

TP→ True Positive: this is positivity consider for the reviews extraction in which 1 represents the true value and 0 represents a false value if the Tp found to be one then all reviews extracted are classify properly.

TN→True Negative: This factor represents the negativity of the results in which it again has two values 0 and 1 if it found to be 1 then the perform test classify wrongly.

Result By Good and Bad Count

Product Name	Total Numbers Of Reviews Extracted	Good Reviews Found	Bad Reviews Found	Better Reviews Found	Best Reviews Found	Word Process	TP	TN
Mobile Samsung	5	4 (a)=5	0 (a)=0	0 (a)=0	1 (a)=1	189	0	1
Tops	5	1 (a)=1	0 (a)=0	1 (a)=1	1 (a)=1	346	1	0
Sport T shirt	4	1 (a)=1	0 (a)=0	1 (a)=1	1 (a)=1	178	1	0
Apple Mobile Phone	5	1 (a)=2	0 (a)=0	1 (a)=1	2 (a)=2	162	0	1
Nokia	6	1 (a)=1	1 (a)=2	2 (a)=1	3 (a)=4	212	0	1

Table 4.1: Result obtained after filtering eviews in Good and Bad

T(p)= Total Product Transaction: =5

This shows the total numbers of product reviews extracted

Recall= TP/Retrieved Document =2/5=0.4

Recall represents the actual accuracy of the execution in which the max value of the recall is not greater than 1. In this recall is calculated by T(p)/Total Retrieved Documents

Result By Probabilistic Approach Implementation

TP (True Positive) =4

Total No of Product (*Retrieved Document*) =5

Recall= True Positive/Retrieved Document =4/5=0.8

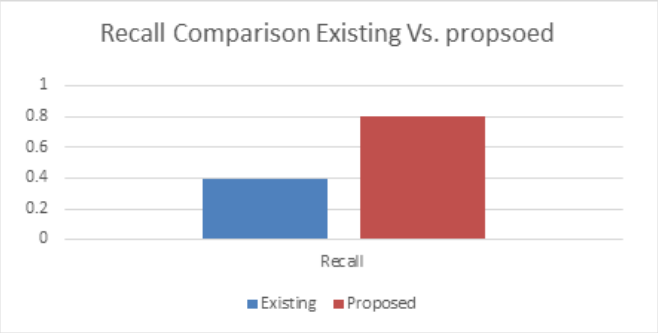


Figure 4.2: Recall comparison

Table 4.2: Result by Probabilistic Approach Implementation

So we can say that probabilistic approach is more convenient and useful to get recall value.

Product Name	Total Numbers Of Reviews Extracted	P (good)	P (Better)	P (Best)	P (Bad)	Word Process	TP	TN
Mobile Samsung	5	0.8	0.0	0.2	0.2	30	1	0
Tops	5	0.2	0.2	0.6	0.0	40	1	0
Sport T shirt	4	0.25	0.25	0.25	0.0	20	1	0
Apple Mobile Phone	5	0.2	0.4	0.2	0.0	96	0	1
Nokia	6	0.166	0.33	0.5	0.33	27	1	0

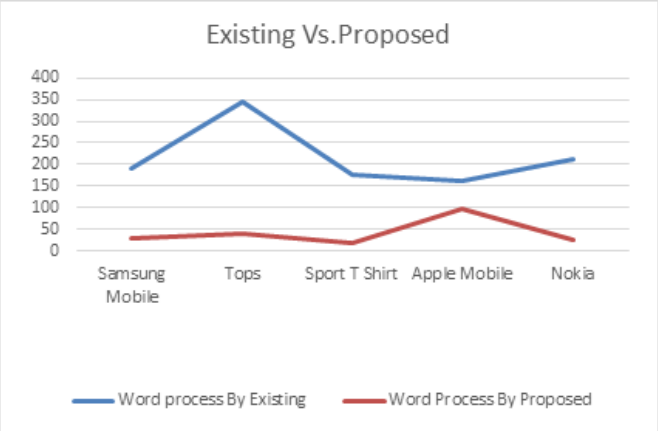


Figure 4.2: Above graph will Represents word process count in all of the products in the categories

V. CONCLUSION

Here we come to conclude that the sentimental analysis used for classification for the reviews to evaluate whether it is positive, negative or neutral reduces the work of manufacturer as well as users. By using aspect-based sentiment analysis user will be able to know about sentiments on feature of product. This will help customer to know about product very well.

VI. REFERENCES

[1]. D. Mali, M. Abhyankar, P. Bhavarathi, K. Gaidhar, M.Bangare International Journal of Management and Applied Science, ISSN: 2394-7926 Volume-2, Issue-1, Jan.-2016

[2]. Magesh Gopu, International Journal of Civil Engineering and Technology , International Journal of Civil Engineering and Technology (IJCIET),Volume 8, Issue 10, October 2017

[3]. Duc-Hong Pham , Anh-Cuong Le, 2018, Data & Knowledge Engineering(2018)

[4]. Anvar Shathik J. & Krishna Prasad K. ,2020,International Journal of Applied Engineering and Management Letters (IJAEML), ISSN: 2581-7000, Vol. 4, No. 2, August 2020.

[5]. Wenhao Zhang, Hua Xu , Wei Wan ,2012Expert Systems with Applications 39 (2012)

[6]. Fu Xianghua , Liu Guo, Guo Yanyan, Wang Zhiqiang, ,2013,Knowledge-Based Systems 37 (2013)

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Using Azure Kubernetes Services to Deploy a Micro Service Based Application

Dr. L. Srinivasan[1], Junaid Ahmed Nadaf[2]
[1]Associate Professor, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India
[2]M. Tech. Scholar, Cyber Forensics and Information Security, ISE Department, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

Micro services are an Independent Services which works perfectly on the architectural approach. Micro services helps in building the applications with its independent behaviour which is well known and also known as small services. This also helps in deploying and managing a service in the form of tiny services where each service is coded separately. The Micro service Architecture is designed in such a way that, if one component fails it won't effect on other running services and ensure that the application services are hosting normally. This Micro service architecture is designed in such a way that the components are loosely coupled and the API's are in well-defined contracts. These benefits in reducing the stress-overhead in the Business Markets. This Micro service feature is offered in one of the open source domain called Kubernetes. This Kubernetes along with Containers and pods helps in deploying, managing and scaling applications called containerized applications. Kubernetes also helps in providing support during any service failure and helps in recovering by improving the situation. In this paper we are going to experiment the Micro services performance with the help of Kubernetes along in its default configuration by enabling High Availability (HA). We had intentionally stopped many services which has helped us to get to know and understand the working mechanism of the same.

Index Terms—Micro services, Containers; Pods, Kubernetes, Failure

I. INTRODUCTION

The Micro services support polyglot programming. And these services communicate with each other by using well- defined APIs. These services are implemented in such a way that they are hidden to each other. The Micro service architecture has management (orchestration) component, which helps in identifying failures, recovering the services across the nodes and deploying the nodes as well. And the API Gateway allows clients to enter inside instead of calling the services, this client when requests for an service, the API redirects to back end and serves the client's request.

The Benefits of Micro services are agility, small teams, and focused teams, small code camp, cross platform technologies, isolating the faults, scalability and data isolation. Some of the Identify applicable funding agency here. If none, delete this.

Major challenges of Micro services are complexity, develop- ment and testing, lack of governance, network congestion and latency, data integrity, management, versioning, and skill set.

Micro services addresses the monolithic approach, which is a drawback. Micro services helps in increase the standard and speed. Micro services always stay low, so that they can easily restart, recover whenever there is service failure. Micro services are known for

reducing the Application downtime value because the services are loosely attached and they don't communicate with each other at all. In order to avail these benefits, there should be technology configured with the existing architecture style.

Containers are known for airy weight and they turn on very quickly than those of Virtual Machines(VMs). Whenever there is a Service Failure, upgrading or in failure cases the containerization of micro services would do its job. Here, we have used Docker, which is mind-blowing container platform. There would be requirement of orchestration platform as well, which helps in deployment of the containers. Kubernetes is an open source technology platform. Apart from healing and restarting the failed containers, Kubernetes has been designed in an automated manner, which identifies the failed states and recovers them automatically. Hence, this feature helps in improvement of application's availability.

The Micro service Architecture style is generic and the practitioners and investigators refers this style as a Model. And with the help of this Architecture style the investigators comes out with different perspectives. After assessment the effectiveness of the High Availability mechanism with the help of Kubernetes, the result logs have been captured. And unfortunately, the results didn't meet the requirement. And the HA effectiveness has been investigated in the Kubernetes in its default configuration mode.

The remaining portion in this paper has been explained in the organized manner as follows. Section II helps us to familiarize Architecture about deploying containerized applications with Kubernetes. Section III explains the experiments which are present. Section IV helps in analysis of the results. And the Conclusion parts have been commented on Section V.

II. IN A PRIVATE CLOUD ENVIRONMENT THE CONTAINERIZED APPLICATION IS DEPLOYED WITH KUBERNETES CLUSTER

It is much needed to know the difference between the Private Cloud and Public Cloud Environment. Kubernetes cluster should be implemented in a public cloud as it requires less effort. Whereas it will take more effort when deploying in Private Cloud. Here the main difference is exposing the Application in a better way externally. We will get to know in details about exposing the applications which is deployed in Kubernetes cluster running in a private cloud. Kubernetes also has the ability to run the Virtual Machines and helps in creating a Single Cluster view. And here the application has been composed of single micro service.

Kubernetes Ingress is a kind of resource which is designed in a structured way to showcase the services. Fig. 1 represents a architecture which is generic design of Ingress which ex- poses the service in a cluster running in a private cloud. A Cluster IP gets generated and helps in redirection of requests to the pods and this Ingress resource is used completely in back end. The Ingress controller always helps in redirecting the incoming requests to the ingress resource, and maps to the appropriate back end service. Only One Pod is used here and Ingress Controller have been deployed here. The rules have to be defined prior on the deployment controller and ensure that the ingress controller has been scheduled on a specific node. The Kubernetes has the possibility to connect to a pod directly to a port on the node which is hosting the pod. We know that we have defined a rule where Ingress controller pod has to be on the same node, and it is safe method where it can always receive the requests from outside of the cluster and that cluster has public IP of that node. Deploying the Ingress Controller and Kubernetes Cluster which is running in the Private Cloud is little tricky and there is no proper documentation available

yet. Understanding and deploying a Ingress Controller needs a lot of trial and errors to understand the working mechanism of the same. The previously described architecture is designed based on our own understanding.

III. EVALUATION RESULTS OF THE MICRO SERVICE BASED APPLICATIONS WHICH IS DEPLOYED WITH KUBERNETES

The nonfunctional requirement which is nothing but Avail- ability is measured usually when there is power outage time from the given period of time. And the Value of High Availability should be calculated when there is a up-time of 99.999 per cent, and we also need to ensure that there should be no power issue more than 5 minutes overall per year. The improvement of the availability of micro services and containers helps in Kubernetes healing capability because they are small and lightweight by natural. So here we are going to describe the experiments which we have conducted to evaluate from an availability perspective and the deployment of the micro service based application in a Kubernetes cluster which is running in a private cloud (Please refer Fig.1).

Whenever there is a Pod Failure, Kubernetes delivers the new pod automatically and therefore this ensures that the improvement of availability of the service is provided by the pod. The evaluation of Kubernetes service failure or pods failure helps in identifying the administrative operations (e.g. pod deletion, node deletion) with the help of Kubernetes command line interface (CLI) and also evaluates how fast Kubernetes delivers the new pods when there is a pod failure. And when there is more usage of Kubernetes administrative operations, where ”failure” is only not to be screened by Kubernetes. Instead, the Kubernetes administrative operations performs it tasks in a graceful manner. Therefore, there won't be any common execution failure scenarios, because there will be a perfect steps followed and these happen spontaneously which is the outcome of Events such as External failure events (e.g. physical node crash). Evaluating the conclusions based on the administrative operations is not accurate. And this is important to know that how to identify and measure the external execution failure and also measuring the availability of the above simulation results before making any decisions.

A. Measuring Availability

From the Availability Perspective, Measuring the Availability helps in evaluating the Kubernetes and some have been defined below.

a) Reaction Time: When there is a failure event detected the time calculated between the introduced failure event and the first reaction of Kubernetes.
b) Repair Time: The time calculated between the repair of the failed pod and the first reaction of Kubernetes.
c) Recovery Time: The time calculated between when the service comes online and the first reaction of Kubernetes.
d) Outage Time / No Power Issue: The time duration in which the service was offline. And hence this Outage Time represents the sum of reaction time and recovery time.

B. The Experiments

The Architecture Diagram can be referred in Fig.1. Here we have deployed three VMs in our Cluster. And this same is hosted in Open Stack Cloud. We have used Open Source OS i.e Ubuntu 16.04 and all the VMs have been deployed here. Container Engine has been introduced by Docker 17.09. And the Main Engine i.e Kubernetes 1.8.2 is providing the nodes which are completely running. Network Time Protocol (NTP) which runs on Port Number 123 is used for time synchronization between the nodes. The Micro service that is officially used here for

experimentation is VLC Video Streaming. The Template used in Pods helps in the deployment controller and this Pod Template also contains the Image of the Containers of the streaming server. And this Container Image when deployed will start to refer the streaming file. And in our Experiment, the main task of deployment controller to maintain only one pod properly amongst all other pods. This also helps in achieving the availability of the Kubernetes Healing Services. And the Video Streaming micro service is emigrant and when there is such a failure, the video streaming service will start to read from the beginning of the file.

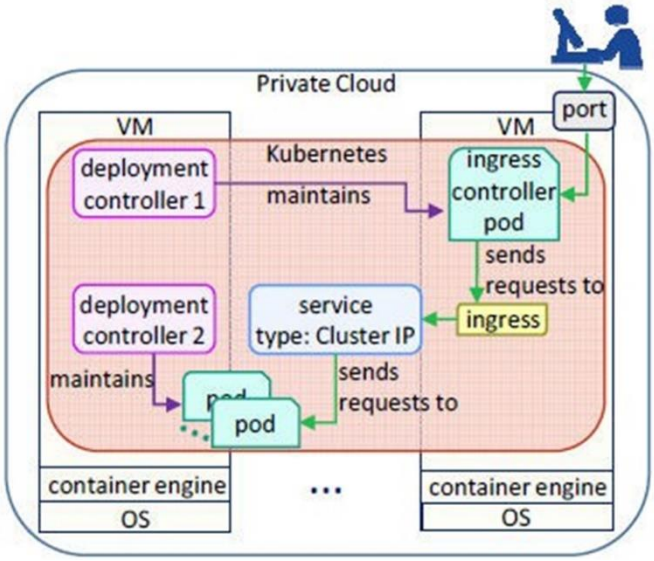


Fig. 1. Private Cloud - Exposing services via Ingress

After successfully passing one round of test, we have identified that there are two sets of Failure Scenarios. The First set is an Application Failure, and this occurs when there is a Pod Failure. The Second set of Failure is known due to Node Failure. In each of these, we will identify and differentiate between these two failure scenarios. Scenario I identifies a Failure reproduced in parallel to an administrative operation which is related to Kubernetes. And this Scenario II is in parallel to external of the Kubernetes. And below we are going to see the Failure scenario of the same in details. And the results of the experiments have been displayed here which represents the failure scenarios and those results can be viewed in Table (I) and Table (II). And all these experiments have been performed say 10 times to get the accurate results. And the results have been filled in the table rows and columns with an average output results taken of those 10 repeated times. And all these Measurements of this experiment have been considered in seconds only.

a) Summary of Pod Failure:

- Plan-I - Pod Failure: With the help of Administrative commands in the CLI the failure pods can be deleted and hence this can also be a showcase to Kubernetes. As explained above that with administrative mode in CLI the failure pods can be removed from the endpoints. And, pods have been given 30 seconds of refined termination period. And at this scenario, the pods won't receive any new requests, and this will also keep serving the requests which was assigned previously. And this also gives an sufficient time to Kubernetes which can be scheduled to a new pod and this also helps in dealing of incoming requests. And the responsible person for maintaining only one replica of pod is deployment controller, and this will bring up the events which always marks the reaction time. And when there is a repair time then the new pod comes up. And this new pod as soon as its landed it starts executing and it starts streaming the video, and this wont be published to users unless and until it has been registered in Endpoints list of the service. And hence, this scenario makes us understand that streaming services are recovered whenever there is a new pod registered on the endpoint list of the service. And the outcome value of this same is presented in Table I.
- Plan-II - Pod Process Failure: Whenever there is a pod which contains some containers i.e Application Con- tainers and when these have

been deployed, parallel to that another container is created which is also the Pod's Container. And these pods are the processes which are running in the OS, may get expelled which ends up resulting in failure of the whole pod. And, to overcome from the pods failure we need to kill the OS Process of the pod container. And the main agent who identifies the failure of Pods process is Kubelet. This Kubelet helps in the reaction time. Similarly, the Kube-proxy helps in removing the pod from the endpoints service list. And when there is a crash in the process, a refined termination signal comes into existence. And this Signal is sent to the application container and the Docker will wait for 30 seconds before being kicked out. And as we have observed in the experiments that whenever there is a Pod Failure and there is Crash over of the pods, the streaming service won't be available. Kubernetes also waits for new pod to start until there will be a termination of the Application Container. simultaneously, the deployment controller will re initiate the pod this is also known as repair time. And the recovery time is calculated when there is a new pod added in the endpoints list of the service. And the measurement of this is represented in Table I.

TABLE I POD FAILURE

Failure Plan (unit: seconds)	Reaction Time	Repair Time	Recovery Time	Outage Time
Plan-I	0.031	0.843	1.422	1.343
Plan-II	0.431	32.443	31.242	32.333

b) Summary of Node Failure:

- Plan-I - Administrative deletion of the Node: A node which gets deleted without draining performed by admin- istrative deletion is one of the method of simulating the node failure. And this node failure is managed by Kuber- netes CLI which is similar to the the Spontaneous Node Failure, and we are going to determine the behaviour of Kubernetes reactions. Here, A node when hosted a pod gets deleted from the Kubernetes CLI Command. And this results in removing the containers and processes which are related to Kubernetes on the active node gets triggered. And the unknown pod which is running on the node which was about to get deleted enters into a state where there wont be any new requests. Therefore the pod failure situation occurs. And we can see lots of different behaviour between this situation and the previous pod failure behaviour(PLAN-1). Also, the pod will serve the requests which were assigned previously just for one second (excluding default 30 seconds). Quickly, the pod is going to be deleted completely and the deployment controller tries to attempt to add a new pod on different node. Here, the repair time gets activated when there is a new born pod. And the recovery time is also taken into consideration when there will be a new pod and this will be added in the endpoint security list. And the calculations for this scenario are represented on Table II.
- Plan II - Externally Triggered Node Failure: As discussed previously, the Kubelet has the responsibility to report the status of the node actively to the master. And in the default configuration of the Kubernetes, the Kubelet gives an update about the Node's Status to the Master Node in every 10 seconds. And this allows the node to be un- responsive for about 4 consecutive status updates before it is confirmed as failed. And after getting confirmed as failed, the master node waits for another 4 minutes and 20 seconds before selecting the pods which are

running on the failed node and which are terminated and those pods are going to be rescheduled on another node. This also means that with the help of default configuration of the Kubernetes, this Kubernetes takes around 5 minutes to be recovered from the node failure. And in order to perform this task, the VM which is hosting the pod is going to be shutdown with the help of Linux Reboot Command.The Master node acknowledges itself that the node is not ready and when there is a missed node status update on the fourth attempt then that would be marked as the reaction time. After 5 minutes, there would be a new pod arrived by the deployment controller, and this pod would be considered and concluded as Repaired. Simultaneously, this new Pod's IP address is added on the endpoints list, and hence the service will be recovered. The assessment value of the same has been represented on the Table II.

IV. ANALYSIS

A. Pod Failure Summary

In the Plan-I case, the reaction time was 0.031 seconds which is very much better than the 0.431 seconds of the failure of pod process(Plan-II). The main reason is that the termination gets activated within the Kubernetes itself, and when there would be a reply appropriately to the termination procedure. And after observing the current state, the Kubelet's health determines that the pod is no longer available and also this helps in predicting how close the next health check failure happens.

TABLE II NODE FAILURE

Failure Plan (unit: seconds)	Reaction Time	Repair Time	Recovery Time	Outage Time
Plan-I	0.031	1.003	1.532	1.513
Plan-II	38.133	25.343	26.432	32.442

The keen observation on the experiments of the same is shown in Fig. 2. The Pod Process gets failed to perform forcefully even after applying the external force which can be read in Plan-II(Fig. 2(b)), and there will be a graceful termination signal when there will be a container of the pod.And hence, the pod process failure gets detected by the Kubelet, and until and unless the Docker gets confirmation from the application container stating that the pod has been terminated. And this leads to the impacts and delays of the SRT(Service Recovery Time) and until the Docker gets new configuration there will be a delay in servicing. And this also ensures that the fault has been occurred due to Bug or might be due to Real Fault. The principles which are required in the Fault isolation has to be responded with proper forceful cleanup of the application containers as soon as there is a confirmation stating that pod's process failure gets detected. And this Grace period gets increased i.e Plan-II is 38.133 seconds. And this is much more significantly longer when there is Plan-I which has Reaction time of 0.031 seconds. Previously, we have observed that in Fig. 2 (a)), Kubernetes always helps in performing the graceful termination and this helps in parallel to repair procedures. For the certain actions of procedures the ordering value is guaranteed for these kinds of procedures. Let's take an example, the terminating pods gets removed from the end points list which helps in recitation of the start of the repair procedures, and the pod gets terminated when there will be a completion and this follows the new pod to be added. This parallelization can be possible only when there will be an assumption that there wont be any fault in the

execution system. This also helps in refreshing that the points that were made earlier about the simulation of the failures which helps on the availability of metrics. And as observed, Plan-I reports an Loss or Outage time of 1.513 seconds and same in Plan-II it reports an loss of 32.44 seconds.

B. Node Failure Summary

And in the node Failure segments, there has been observed that there is an similar differences in all the framed availability output. And for Plan-I, the failure has been triggered from the Shell of Kubernetes, there will be a Reaction time of 0.031 seconds and it is much faster than that of the Reaction time captured in Plan-II i.e 38.133 seconds. As explained in the previously section that, the period of Kubelet's status depends on the period which gets updated by default in 10 seconds and the same is allowed in the number of Missed status by default in 4 seconds.

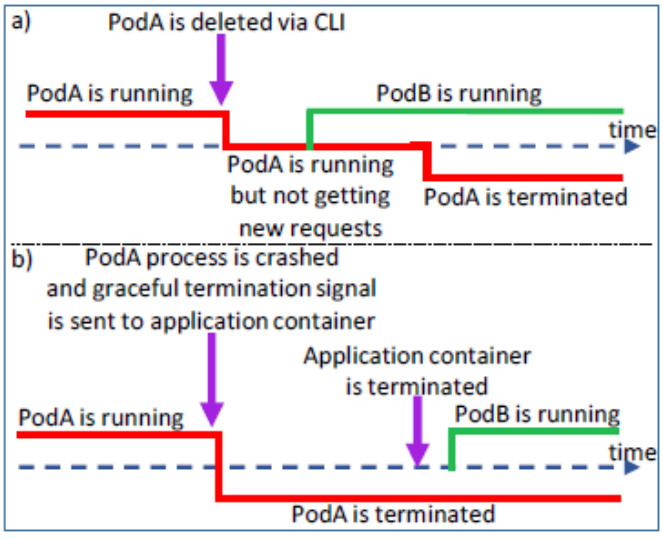


Fig. 2. Analysis of Pod Failures Plans. (a) Plan-I (b) Plan-II

And another important point which was observed in Plan-I is that, though the pod has failed to perform from the Kubernetes, the new pod is going to be initiated and the old pod is going to be terminated. And hence this was expected to behave properly similarly to that of the termination of the pods by administrative method(Plan-I), and hence the new pod will be procured prior to the old one which gets terminated.

V. CONCLUSION

Kubernetes helps in enabling the healing service whenever there is a recovery failure. And this is evaluated through internal operations. Hence, for these kinds of operations, Kubernetes helps to react will in comparison with its reaction to failures which results in performing the external triggers. And according to the survey experiment, the downtime of the Kubernetes service is significantly higher. And it is much important to know that the default configuration of the Ku- bernetes leads to trigger the node failure forcefully. And the output tables explains that these types of failures leads to downtime for 5 minutes, and this in results helps in the amount of downtime calculation for one-complete year and for the high availability requirements are not differentiated and these do not gets satisfied automatically even if an application or a micro service gets deployed in Kubernetes.

As we know that the default configuration can be altered in Kubernetes, analyzing how to reconfigure the Kubernetes and how the reaction would be when there is node failure. This helps in avoiding the network overhead and there would be fake reports which makes it complicated and which requires a better effort.

VI. ACKNOWLEDGMENT

This complete work has been demonstrated and implemented from the NSERC and Ericsson.

VII. REFERENCES

[1]. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2]. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol.2. Oxford: Clarendon, 1892, pp.68–73.
[3]. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4]. K. Elissa, “Title of paper if known,” unpublished.
[5]. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[6]. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7]. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: Univer- sity Science, 1989.
[8]. S. Newman, “Building Microservices: Designing Fine-Grained Sys- tems,” O’Reilly Media, Inc., 2015.
[9]. N. Dragoni et al., “Microservices: Yesterday, Today, and Tomorrow,”in Present and Ulterior Software Engineering, Springer, Cham, 2017, pp. 195–216.
[10]. D. Jaramillo, D. V. Nguyen, and R. Smart, “Leveraging microservices architecture by using Docker technology,” in SoutheastCon 2016, 2016, pp. 1–5.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Prediction of Car Price When Using a Machine Learning Technique

Prof. Pradip Sitaram Ingle[1], Mr. Akshay Shivaji Sonune[2], Mr. Vaibhav Subhash Borade[2]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

A car price prediction has been high Area of interest research, as it requires noticeable effort and knowledge from the field expert. Considerable Number of different attributes are examined for Reliable and accurate prediction. To build a model for Predict the price of used cars in Bosnia and In Herzegovina we used three machine learning Techniques (Artificial Neural Network, Support Vector Machine and random forest). However, the Mentioned techniques were used to function as one Ensemble. The data used for the prediction were collected from the web portal autopijaca.ba using the web Scrape that was written in PHP programming Language. Respective performances of different Algorithms were then compared to find one that was best Fits available data set. The final prediction model was integrated into the Java application. Furthermore, it is the model was evaluated using test data and accuracy of 87.38% was obtained.

Keywords – Car Prediction, Support Vector Machines, Classification, Machine Learning.

I. INTRODUCTION

This paper Predict the price of used cars in both an important and interesting issue. According to data obtained from the National Transport Authority [1], the number of registered cars between 2003 and 2013 has witnessed a spectacular increase of 234%. With difficult economic conditions, it is likely that sales of used imported (reconditioned) cars and used cars will increase. It is reported in [2] that the sale of Prediction According to information received From the Agency for Statistics of BiH, 921,456Vehicles were registered in 2014, of which 84% ofThey are cars for personal use [1]. This number is increased by 2.7% since 2013, and it is likely that this The trend will continue, and the number of cars will Accurate car prediction involves expert Knowledge, because price usually depends on many Characteristics and factors. Usually most Important are make and model, age,ss Horsepower and mileage. The type of fuel used in Car and fuel consumption per kilometer have a big impact Price of a car due to frequent changes in the price of A fuel. Various features such as exterior color, door Number, type of transmission, dimensions, safety, air Condition, interior, whether it has navigation or not Will also affect the car price. In this article, we Applied various methods and techniques to Achieve higher precision of the price of used car Prediction. Accurate car prediction involves expert Knowledge, because price usually depends on many Characteristics and factors. Usually most Important are make and model, age Horsepower and mileage. The type of fuel used in Car and fuel consumption per kilometer have a big impact Price of

a car due to frequent changes in the price of A fuel. Various features such as exterior color, door Number, type of transmission, dimensions, safety, air Condition, interior, whether it has navigation or not Will also affect the car price. In this article, we Applied various methods and techniques to Achieve higher precision of the price of used car Prediction This paper is organized as follows: Section II contains related work in the price field Predictions of used cars. In Part III, the research The methodology for our study is explain. Section IV Elaborates on various machine learning algorithms and Examine their respective perceptions to predict The price of used cars. Finally, in section V, a The end of our work is given, along with Future work schedule.

II. RELATED WORK

Surprisingly, the work with the estimated price of used cars is very recent, but also very sparse. In his MSc thesis [3], Listiani showed that regression mode building using support vector machines (SVM) can estimate the residual price of rental cars with higher accuracy than simple multiple regression or multivariate regression. SVM is Predict the price of used cars using machine learning techniques

Predict the price of used cars has been studied Extensive in various surveys. Listian discussed, In her thesis written for master's thesis [2], thatRegression model built using Support Vector Machines (SVM) can predict the price of a car like Has been leased with better precision than Multivariate regression or some simple multiple Regression. This is on the grounds that Support Vector Machine (SVM) is better at handling datasets With more dimensions and it is less prone to Top fittings and under equipment. The weakness of this Research is that a change of simple regression with More advanced SVM regression was not shown in Basic indicators such as mean, variance or standard

III. DEVIATION

As average deviation or deviation. In another university thesis [4], Richardson worked with the hypothesis that car manufacturers are more willing to produce vehicles that do not weaken quickly. In particular, using a multiple regression analysis, he showed that hybrid cars (cars that use two different power sources to power the car, ie they have both an internal combustion engine and an electric motor) are more able to retain their value than traditional ones. Vehicle. This is probably due to more environmental considerations for the climate and due to higher fuel efficiency. The importance of other factors such as age, mileage, brand and MPG (miles per gallon) were also considered in this study. He collected all his data from various websites.

Table 2. Minimum and Maximum Values

#	CYLINDER VOLUME (CC)	YEAR	PRICE (RS)
Minimum	1000	1988	27, 000
Maximum	2160	2013	825, 000

Wu et al. [5] used neuro-fuzzy knowledge-based system to predict the price of used cars. Only three factors, namely: the car brand, the year it was produced and the engine style were considered in this study. The proposed system gave similar results compared to simple regression methods. Car dealers in the United States sell hundreds of thousands of cars each year through leasing [6]. Selling these cars at the right price has great financial significance for their success. In response to this, the ODAV system (Optimal Distribution of Auction Vehicles) was developed by Du et al. [6]. This system not only estimates the best price for the resale of the cars, but also gives advice on where to sell the car. Since the United States is a large country, the place where the car is sold also has a non-trivial impact on the selling price of used cars. A k-nearest neighbor regression model was used to forecast the price. Since this system was launched in 2003, more than two million vehicles have been distributed through this system

[6]. Gonggi [7] proposed a new model based on artificial neural networks to predict the residual value of private used cars. The main features used in this study were: mileage, manufacturer and estimated service life. The model was optimized to handle non-linear conditions that cannot be done with simple linear regression methods. 3. Methodology Data were collected from << petites announce >> found in daily newspapers such as L'Express [8] and Le In the related work shown above, authors Proposed prediction model based on the single Machine learning algorithm. However, it is noticeable The approach to machine learning algorithms did Do not give remarkable predictions and can be Improved by collecting different machine learning Methods in an ensemble

IV. METHODOLOGY

Data were collected from << petites announce >> found in daily newspapers such as L'Express [8] and Le Defi [9]. We made sure that all the data was collected in less than a month interval, as the time itself could have a noticeable impact on the price of cars. In Mauritius, seasonal patterns are not really an issue, as this does not really affect the purchase or sale of cars. The following data was collected for each car: make, model, cylinder volume (funny, it is considered the same as horsepower in Mauritius), mileage in km, year of manufacture, paint color, manual / automatic and price. Only cars that had their price listed were registered.

Thus, paint color and manual / automatic functions were removed. The data were then further adjusted to remove records where either age (years) or cylinder volume were not available. The model was also removed, as it would have been extremely difficult to get enough records for all the different car models that exist. Although the mileage data were sparse, they were retained as they are considered a

A selection of collected data is shown below in the table

Table 1. Sample Data Collection

#	MAKE	CYLINDER VOLUME (CC)	YEAR	MILEAGE/KM	PRICE (RS)
1	TOYOTA	1300	2007	38000	410000
2	NISSAN	1500	2007	50000	325000
3	HONDA	1500	2005	59000	385000
4	TOYOTA	1000	2007	59000	360000
5	TOYOTA	1300	1989	62665	50000
6	TOYOTA	1500	2008	67000	615000
7	TOYOTA	1500	2008	69000	575000
8	TOYOTA	1490	2006	73000	450000
9	TOYOTA	1600	2006	82000	550000
10	TOYOTA	1000	2006	85000	325000
11	TOYOTA	1500	2000	113000	325000
12	TOYOTA	1500	2000	129000	218000
13	NISSAN	1500	2001	145000	195000

Initially, 400+ records were collected. After further pruning, for example, we kept only the three of the most popular brands in Mauritius, ie Toyota, Nissan and Honda. In particular, we removed all brands for which there were less than ten records. In terms of cylinder volume, for some cars it was given in a range. We then selected the average value for the site. The values are then pretreated in a form suitable for further processing using machine learning techniques. The minimum and maximum values for some numeric functions are shown in Table 2.

The collected raw data set contains 1105 samples. SinceSince data is collected using a web scraper, it is. Many examples that have only a few attributes. In order To clean these samples, PHP script reader Deleted data from the database, perform cleaning and Saves the cleaned samples in the CSV file. CSV The file is later used to load data into WEKA software To build machine learning models [10]. After the cleanup process, the dataset has been Reduced to 797 samples. Especially all brands like Have less than 10 samples and where the price isHigher than 60,000 BAM was removed due to Skew class problem.

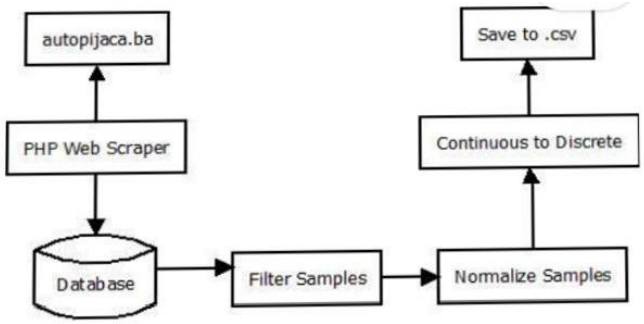


Figure 2. Data gathering and transformation workflow diagram

entire data creation process is displayed in Figure 2.

V. CONCLUSION

Predicting car prices can be a challenging task To the high number of attributes to be Considered for accurate prediction. The big step In the divination process is gathering and Pre-processing of the data. In this survey, PHP Scripts were built to normalize, standardize and clean Data to avoid unnecessary noise for machine learning Algorithms. Data cleaning is one of the processes that Increases prediction performance, but still insufficient for The cases of complex datasets like the one in this Surveys. Use of algorithm for one machine on The accuracy of the dataset was less than 50%. Therefore Ensemble of several machine learning algorithms Has been proposed and this combination of ML Methods get an accuracy of 92.38%. This is Significant improvement over a single machine Learning method approach. However, the downside Of the proposed system is that it uses a lot More computational resources than one machine Learning algorithm. Although this system has achieved amazing results Performance in the car price prediction problem our goal For future research is to test this system to work Successful with different datasets

VI. REFERENCES

[1]. Agencija za statistiku BiH. (n.d.), retrieved From: http://www.bhas.ba . [accessed July 18, 2018.]

[2]. Listiani, M. (2009). Support vector regression Analysis for price prediction in car rental Application (doctoral dissertation, master's thesis,TU Hamburg-Harburg).

[3]. Richardson, M. S. (2009). Determinants of used car Resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object/ coccc% 3A1346 [opened: August 1, 2018.]

[4]. Wu, J. D., Hsu, C. C., and Chen, H. C. (2009). An Expert price warning system for used cars Using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36 (4), 7809-7817.

[5]. Du, J., Xie, L., & Schroeder, S. (2009). Practice Price paper – PIN optimal distribution of auction Vehicles System: Use of price warning, Estimation of elasticity and genetic algorithms for Distribution of used vehicles. Marketing Science, 28 (4), 637-644.

[6]. Gongqi, S., Yansong, W., and Qiang, Z. (2011,January). New model for prediction values Of the used car based on the BP Neural Network and Non-linear curve fitting. In measurement technology and Mechatronics Automation (ICMTMA), 2011 third International Conference on (Vol. 2, pp. 682-685). IEEE.

[7]. Pudaruth, S. (2014). Predict the price of used cars Using machine learning techniques. Int. J. Inf. Calculate. Technol, 4 (7), 753-764.

[8]. Noor, K., & Jan, S. (2017). Prediction for vehicles System using machine learning Techniques. International Journal of Computer Applications, 167 (9), 27-31.

[9]. BiH Auto Pijaca. (n.d.), Retrieved From: https://www.autopijaca.ba. [opened August 10, 2018].

[10]. Week 3 – Data Mining with Open Source Machine Learning software in Java. (n.d.),

Retrieved From: https://www.cs.waikato.ac.nz/ml/weka/.[4. August 2018].

[11]. Ho, T. K. (1995, August). Random decision forests. In Document Analysis and Recognition, 1995., Negotiations for the third international conference On (Vol. 1, pp. 278-282). IEEE.

[12]. Russell, S. (2015). Artificial intelligence: A modern Approach (3rd edition). PE.

[13]. Ben-Hur, A., Horn, D., Siegelmann, H. T. and Vapnik, V. (2001). Support vector clustering. Journal of Machine learning research, 2 (Dec), 125-137.

[14]. Aizerman, M. A. (1964). Theoretical basis for The potential function method in pattern recognition Learning. Automation and remote control, 25, 821-837.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

An Intelligent Trash can use Wireless Sensor Network Technology

Prof. Pradip Sitaram Ingle[1], Ms. Shital Prabhu Gadkari[2], Ms. Pranita Shivkumar Murkar[3]

[1]Assistant Professor, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India

[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

Planning of the rubbish bin of reprocessing authority of all rubbish formed in community Assistant is a big responsibility . The collection of unwanted items and repressing or waste authority contributes to community neatness and public health. Each people must helps for the conservation of any resources building authority office needs a batter administration structure that has the right collection plan garbage in different location and when need. The focusing structure can ensure the freshness of the community it gives efficient way to any trash collector. This paper Gives the solution to the situation that is an intelligent Control structure for waste reprocessing using wireless sensor Network technology that can be Technology used in all elevated buildings

I. INTRODUCTION

WSN is the group of specialized sensor and transducer . the large number of sensor nodes is present in the WSN is nodes smaller in size as compare to other nodes in current networks which sense the actual time events processing . Increase in population , a change in living lifestyle and an increased number of industries, the amount of municipal waste (MSW) is increasing at a very high rate [1]. Currently, the amount of solid waste produced in urban India is 68.8 million tones per year [2]. This amount is expected to double by 2025 [1]. This shows the need for the right waste management solutions so that the harmful effects on the environment can be reduced.

The waste management cyclic system includes the generation of waste from industries, houses, markets, etc. as the waste is thrown in the rubbish bins. This waste is further picked up by the municipal companies to finally dump it in dumping areas and landfills. However, due to a lack of resources, inefficient groundwork, no waste is c in this field.

Proper cleaning interval can provide a solution to this problem. But keeping track of garbage status manually is a very difficult job. An answer to this problem is proposed in this article in the form of a framework that uses a wireless sensor . Related work,

proposed system framework, results and discussion and conclusion and future work have been discussed in the following sections to provide detailed information about the work done Environment, plays an agency that handles waste a central role in any community.

II. RELATED WORK

An RFID and a load cell sensor-based waste management system are implemented in paper [3]. The parameter of garbage monitored in this system is the amount of waste dumped in the garbage. The collection interval is determined on the basis of this parameter. The use of authentication passwords based on RFID technology helps to provide extra security in this system by identifying stolen junk. At a time Many of the earlier functions were developed, such as automated recycling systems. With new tracking technology, the focus is on separating waste as aluminum, glass and plastic containers for waste management. It also uses radio frequency identifier or Bluetooth low energy technology to facilitate. In off-waste management, smart systems provide the right to handle waste efficiently, helping people to recycle it in the process. [11] Theoretical framework and algorithm are developed in this article for successful implementation of hardware. The retrieved information is stored for monitoring and management of activities.

Another framework based on RFID, GSM and GIS is proposed in paper . The proposed system monitors the waste management and management process. This system provides real-time data for the waste collection process, tracking of the vehicle's position is done using GIS which helps to overcome difficulties such as route optimization.

An application based on distributed sensor technology and geographic information system to be used for monitoring municipal solid waste has been proposed in paper . A case study based in the Pudong area is presented in this article. The most important outcome of this article is the calculation of waste weight and volume to be used further to optimize routes for garbage collection vehicles and material density assessment. Rubbish that can detect if the rubbish bin gets crowded by placing one

Photoelectric sensor. The trash can also find the category Waste tossed in the trash and squeeze the trash if it is Overflow. The measurement is sent through Knowledge is received by the RFID reader

In Somu Dhana Satyamanikant, a smart rubbish has been designed to detect whether the trash can is crowded with photoelectric sensors. The weight sensors used in it fall into it Used to measure the weight of waste waste. Measurement information is sent for weighing, it is received by the RFID reader. And the garbage dumped in it can also detect the type of garbage dumped in it and squeeze it out if the garbage overflows.

A new architecture focused on senser nodes that uses to transfer information from the trash cans to a remote server has been implemented by some researchers in paper [10]. In this framework, a single

parameter is monitored and the value is stored on a remote server that provides a web server to interact with the user. The author uses Argos fashion in this paper which only covers 430m geographical area.

A Zigbee Pro and GPRS focused on waste authority stature has been proposed in application .The structure framework monitors the amount of waste by load , climate and heaviness in the rubbish box, and residual power in the rubbish bin and updates the data to a control station. The final data stored in the form of a database is presented at the end of the assignment.

The proposed framework is a three tier system named as 1) smart Box , 2) arch 3) Remote Base station

Garbage collection systems are very functional, such as optimization of selected waste bins. Uses to find effective root energy using Android application.

Real-time monitoring using the Google Maps API used to read, monitor, and migrate junk levels such as trashed garbage [18, 24] and garbage weight and height. Optimization of selected rubbish bin [17], Use of clone function to find the efficient way to Filled rubbish bin [18, 24] and actual time mortaring using Google map for seen , contorting and parasling junk levels Essentially load and elevations of the trash .

TABLE 1 encapsulated certain necessary systems for reprocess waste, essentially

As together just as necessary functions.

Theatre some complications in these linked works [8, 11, 14]

• Overflow condition that cannot be handled effectively

The cleaning agents.

• There is no proper plan for detergents to pick up the trash

From the trash.

• A lot of time wasted with unplanned trips at different levels

To ensure waste management.

• Inefficient workforce and time management leading to Inefficient cost.

III. METHODOLOGY

The design proposed here is like any other developed system, But it does provide information about the trash can Android function . Category of structure is assembled from Questionnaire taken a way are required for huge-rise buildings trig trdential apartments. The design process starts with drawing Smart Recycling system and architecture diagram for System thats uses Firebase platform to communicate Inserted two devise, a micro controller along sensors and enjoyer Along the structure datasets . The function To discuss with firebase, Android Studio is pre-owned When implementing design, ultrasonic sensors such asLocated on top of the trash can, the interface is also includedNode MCU Arduino as shown in Figure 2. Node MCU is

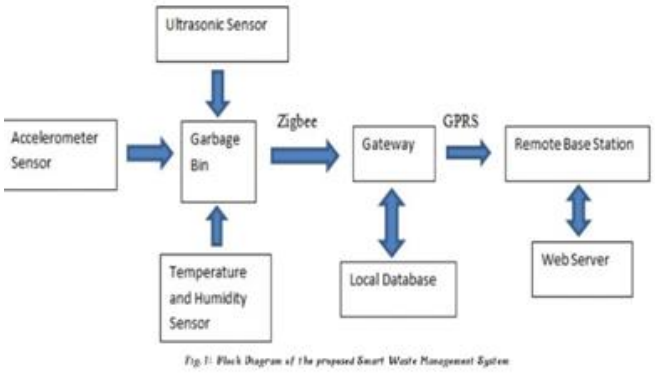


Fig.1: Block Diagram of the proposed Smart Waste Management System

Programmed using the Arduino to tack and convey Necessary knowledge given from the trash, although Android The application is formed by using Android Studio. To Send required alerts to the Android application from Free base Slave , Free_base Cloud Messaging (FCM) Alerting supply is used when programming languages Used in this function are React Native and JAVA [22].

Auther's	Methods	Features	Performance Evaluation
[7]	Zigbee and Global System for Mobile Communication (GSM)	The main feature of this work is a text is sent to garbage truck driver.	This method only works only for the truck cleaner
[8]	using a RFID-based system integrating the web-based information system at the host server	The system can classify materials and separate waste	Able to classify the recycled items but it is not able to send report to the rest of the recycling bins
[10]	Applications of RFID technology	Product self-management, with emphasis on municipal solid waste management as well as environmental implications of RFID	This application is only for solid waste detection, separation and collection three materials such as the aluminum, plastic, and glass containers,
[12]	Microcontroller with the integration of sensors and mechanisms enable effective recognition and automatic separation of recycled items.	it will feature the integration between microcontrollers, sensors, stored value card, programming and mechanical mechanism	The method only works similar with Reverse Vending Machines, it has a casing, which has 3 parts for detection only.

Smarty Reprocess Bin System was approved using Various test cases in this phase. Crib cases were prepared to Test all paths available in the codes of NodeMCU and Android application. three types of testing; entity Examination, assimilation examination and structure examination. You were Performed to sure about the reusability of the structure. Various devices of Modern reprocessing stacture was examined uniquely in the entity examination while the entire system was examined at the time of assimilation and structure examination[20,23].

IV. RThe database that contains the values of various parameters in the trash is discussed in this section. The data from the garbage is collected for 5 consecutive days. The database is created online using a program called Caspio. The login page for the database . The name in the database is Smart Waste Management System. Creation date and last change date are mentioned on the login page. The various options such as adding tables, data pages, adding authentication of stored data, changing the style of the database, etc. are mentioned on the left side of the login page. Hardware's and programming application are one of the smart trash The statured is completely examined. Waste level details like its The elevation is monitored. The plase of the rubbish is also show in The central assistant. Pattern and appliances Exertion of smart wastage.

IV. CONCLUSION

This article proposes a smart waste management system that focuses on the waste management process. The status of the checkout is continuously monitored at the control station and presented in a

graphical user interface to provide a user interaction with the system. The values stored in the database help a user to have the updated data in the box, as well as the previous values of the parameters in the box. This collected data from the garbage can be used to optimize routes for the collection cars for efficient use of resources in the suggested devastation administration structure With the increase in meccanization and populace,

Sturdy waste administration has comes a big problem. This modern Reprocessing Bin Stricture gives a overall solution This structure offers actual time Garbage disposal control stricture and it uses in main way.

The structure for confirm the konwlege required for waste Authority people. By joining the structure to the mobile Applications, it helps the garbage collector cleaners to accept Prise-effitient konwlege. The datasets helps the report to be Generated. Endly, the project has been successfully completed And archive all the goals.

V. REFERENCES

[1]. Saeed, M. O., Hassan, M. N., & Mujeebu, M. A. (2009). Evaluation Of municipal waste production and potential for recyclable material In Kuala Lumpur, Malaysia. Waste Management, 29 (7), 2209-2213.

[2]. D. Hoornweg and P. Bhada-Tata, World Banks: What a Waste – A Global Review of Solid Waste Management, Urban Development & Local Government Unit, World Bank, 1818 H Street, NW, Washington, DC 20433 USA, 2012.

[3]. http://thekachraproject.in/infographic-m kommunal-solid-waste-msw-in-india / opened March 2, 2016.

[4]. howdhury, B., & Chowdhury, M. U. RFID based Real Time Smart Waste Management plan in Telecommunication Networks and Applications Conference (ATNAC), IEEE, pp. 175-180, 2007

[5]. Hannan, MA, Arebey, M., Begum, RA, & Basri, H. Radio Frequency Identification and Communication Technologies for Solid Waste Bin and Truck Monitoring System' in Waste management, Vol. 31, edition 12, p. 2406-2413, 2011

[6]. Rovetta, A., Xiumin, F., Vicentini, F., Minghua, Z., Giusti, A., & Qichang, H. Early Detection and Evaluation of Waste through Sensoriz Containers for a Collection Monitoring Application in Waste Management, Vol. 29, edition 12, pp. 2939-2949, 2009

[7]. Faccio, M., Persona, A., & Zanin, G. Waste collection many objective model with real time traceability data' in Waste Management.

Second National Conference on Internet of Things : Solution for Societal Needs
In association with International Journal of Scientific Research in Computer Science, Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Waste Management System for Automatic Alert of Filling of Dustbins Using IOT

Dr. Avinash S. Kapse[1], Ms. Vaishnavi S. Narkhede[2], Ms. Poonam Wankhede[2], Ms. Nutan Bhanuse[2], Ms. Raksha Arakh[2]

[1]HOD, Department of Information Technology, Anuradha Engineering College Chikhli, Maharashtra, India
[2]Student, Information Technology Department, Anuradha Engineering College, Chikhli, Maharashtra, India

ABSTRACT

National campaign with Government of India, closed. The aim is to cover all the rural and urban areas of the world so that the country becomes international class with many Internet of Things (IoT) devices such as Smartphones & sensors. One of the biggest problems with our society is the power problem. Waste detection, inspection and management is one of the first issues of this era.

Automated manufacturing methods are used in almost all major life forms. Solid waste is a source and Cause of environmental pollution defined under the Resource Conservation and Recovery Act as all solid liquids, semi-solids or gum materials disposed of from industrial, commercial, mining or agricultural activities and from Community events. Often in our city we find trash cans or trash cans placed in public places. To avoid all these situations, we will be installing the Intelligent Waste Recovery Project on IoT and waste collection. This trash can is compatible with microcontroller-based systems with ultrasonic sensor systems as well as centralized systems Generalize the current status of the trash, in mobile browsers with html pages and Wi-Fi. Hence, the position will remain on the html page. Generally our projects depend on the functionality of these Wi-Fi modules; important for its implementation.

In the current waste management system, local governments manage waste by installing bins and operating several waste collection companies. is a price structure using the same set of costs, which causes environmental problems and increases Storm due local in governments manage waste by installing bins I manage many waste management companies. Because the price of using the same cost to cause environmental problems and storms has risen, because there are no restrictions for large food producers and no incentives for simple producers. To address this issue with the current waste analysis, an IoT-based waste management system has been launched. Internet of Things (IoT) is a term used everywhere to connect wired and wireless networks without user intervention.

I. INTRODUCTION

India's growing population poses serious threats to the availability of living space, the use of natural Resources and raw materials, education and employment. But another serious danger that follows is the increase in the number of Waste generated every minute by an individual. Every city faces the threat of ever increasing waste. A

1.1 million tons of waste are generated every day in our country. Unfortunately, only 5% of this colossal amount of waste

1.2 is recycled. In India, the collection, transport and disposal of MSDs is unscientific and chaotic. Uncontrolled spill

waste from urban neighborhoods has created overflowing landfills that are not only impossible to recover due to a disorderly spill, but also have serious impacts on the environment. On a large scale, the low recovery rate has hampered the country's growth and the country's economy. We arrived at an automatic waste separator, which classifies wet waste as dry. This project is a highly innovative IoT waste control system that will help keep the city clean. The system controls the bins and reports the level of waste collected in the bins through the Android app. It should be controlled by the administrator. current waste management system,

1.1 Motivation

Clean India Mission is a national campaign of the Indian government, covering towns and to clean up villages, country streets, roads and infrastructure.

Choosing this topic is intended to help us in this mission through our engineering Knowledge for the benefit of our society.

A significant increase in the production of municipal solid waste has been recorded in the separation of waste worldwide. Effective management must be implemented in order to live a better planet.

Therefore, we try to make a change with the cost-effective project proposal. It is responsible for minimizing the use of the blue collar method of disposing of waste in an automated panache. Year Automation in this style not only preserves the manual separators of many health problems, but also proves that it is economical for the nation. In addition, this system uses low-cost components for successful separation of most types. It is also becoming an important issue in academia, industry and government as an important area of IoT application.

II. LITERATURE SURVEY

"Smart garbage collection system in a residential area" - In this project, solid waste management in urban areas is a major challenge for most countries around the world. Effective waste management is a prerequisite for maintaining a safe green environment as waste disposal of all types is increasing. Bibliography survey 1 "Intelligent garbage collection system in a residential area" - In this project, solid waste management is largely a major challenge In urban areas in most countries of the world. Effective waste management is a prerequisite for maintaining a safe green environment as more and more all types of waste are disposed of. The main concept is that camera A be installed at each garbage collection point with a load cell sensor at the bottom of the garbage. Tea will take the camera continuously snapshots of the trash. Many technologies are used for waste collection as well as for well-managed recycling. Information The gathering is large and heavy. The concomitant effects of a rapid national growth rate, a large and dense residential stock And a pressing demand for the protection of the urban environment creates a difficult framework for waste management. The complexity of the context and procedures due to problems is a major concern for local city authorities Today's municipal solid waste collection, transportation and treatment involves manual and time- consuming collection. If it is installed in apartments or small colonies, it is advantageous to collect separately at the landfill site. This is the purpose of our project. Because of the characteristics and advantages of IoT services, a separate collection bin is using a smart bin "- What our project aims to make a municipal waste management system useful at home. The Categories in which the waste is composted are dry, wet, metallic. Continued increase in population increases waste production. As production and consumption

multiply, a significant amount of solid matter and human release are regularly generated. Mountains of waste are common today. Landfill waste is omnipotent in the form of rotten piles that dot our land and make our rivers, wells and lakes hateful. India generates 68.8 million tonnes of municipal solid waste every year. When collected, waste that cannot be used outside is released, resulting in leachate and gaseous emissions. Pollution of the surrounding environment. Since this system aims to deliver results at the household level, its goal is municipal solid waste (MSW). Waste is divided into three categories dry, wet and metallic. Wet waste the household level can be Vegetable peels, garden leaves, weeds, dried fruits, etc.

III. PROPOSED METHODOLOGY

The block diagram consists of Arduino Uno components, fC-28 humidity sensors, HC-SR04 ultrasonic sensors, DC motor, relay and ESP8266 Wi-Fi modules.

• A humidity sensor is used to detect whether the waste is dry or wet.

• Two DC motors are used; one is to move the conveyor belt and the other is to place the rotary trash can to collect the garbage in another bin.

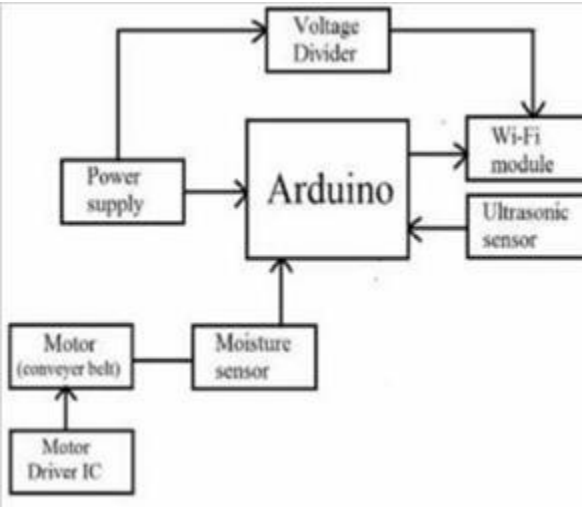


Fig.1 Block diagram

ESP8266 Wi-Fi module is used to update the status of Trash cans on the mobile application.

Fig.2 Data Flow Diagram

Relays are used to drive DC motors.

• One is used to detect the level of dry waste and the second to detect the level of waste in the wet container.

The sensor will detect if the garbage is dry or wet, if the garbage is dry, it is collected in the dry part of the trash and if the garbage is wet then the garbage can move 180 degrees and collect the garbage from the wet side. trash can.

• Arduino then sends this information to the wifi module and the wifi module updates this information in the mobile application.

• The two ultrasonic sensors are then connected to the VCC and GND Arduino 5v and GND.

VCC & GND Both of the relays are connected to VCC and GND in Arduino.

• Now download the Arduino hex file, after downloading the hex file, press the RUN button. The virtual terminal will display the distance

The measurement, that is, the basket is full or empty. The images below show the mobile application and the state of the trash in real time, as the database is 50% dry and 100% wet. A Wi-Fi module pin connected to the Arduino reset pin.

• The result

• Here we use a variable voltage source and set -250V as the threshold value. If we change the voltage below the threshold value, we will not fill it in the virtual terminal that is the trash.

IV. CONCLUSION

In Proteus, we connect this variable voltage source to the analog pin of the ultrasonic sensor, connect the trigger and echo to the Arduino, and use the potentiometer as a humidity sensor and connect the Arduino to the A0 pin.

V. REFERENCES

[1]. Tarandeep Singh , Rita Mahajan , Deepak Bagai, "Smart Waste Management using Wireless Sensor Network", in IJRCCE Volume 4 , Issue 6 , June 2016.

[2]. Collection Bin", IJARECE) Volume 5, Issue 5, May 2016.

[3]. Prakash, Prabu," IoT Based Waste Management for Smart City", published in IJRCCE Volume 4, Issue 2, February 2016.

[4]. Narayan Sharma, "Smart Bin Implemented for Smart City",I nternational Journal of Scientific & Engineering Research, Volume 6, Issue 9, September-2015

[5]. KanchanMahajan, "Waste Bin Monitoring System UsingIntegrated T echnologies", International Journal of Innovative Research in Science, Engineering and Technology, Issue 3, Issue 7, July 2014.

[6]. Issac, R;Akshai,M. "An effective solid waste management system for Thiruvalla Municipality in Android OS" IEEE Conference Publications , 2013.

[7]. M. Al-Maaded, N. K. Madi, RamazanKahraman, A. Hodzic, N. G. Ozerkan , An Overview of Solid Waste

Second National Conference on Internet of Things : Solution for Societal Needs

Organized by

Department of Information Science and Engineering,
New Horizon College of Engineering,
Kaverappa Layout,Near Marathahalli,
Bengaluru-560103 Karnataka, India

Publisher

Technoscience Academy

Website : www.technoscienceacademy.com
Email: info@technoscienceacademy.com

Email : editor@ijsrcseit.com, ijsrcseit@gmail.com